DOCUMENT RESUME

ED 132 192 95 TM 005 921

AUTHOR Coles, Gary J.; Chalupsky, Albert B.
TITLE Impact of Educational Innovation on Student Performance: Overall Findings for Reading and Arithmetic. Project LONGSTEP Final Report: Volume I. Supplement.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO AIR-21400-9/76-FR-SUPP
PUB DATE Sep 76
CONTRACT OEC-0-70-4789
NOTE 157p.; For related documents, see TM 005 891-896 and TM 005 920-922

EDRS PRICE MF-$0.83 HC-$8.69 Plus Postage.
DESCRIPTORS *Academic Achievement; *Arithmetic; *Educational Innovation; Elementary School Students; Elementary Secondary Education; *Longitudinal Studies; *Reading Achievement
IDENTIFIERS Longitudinal Study of Educational Practices; *Project LONGSTEP

ABSTRACT

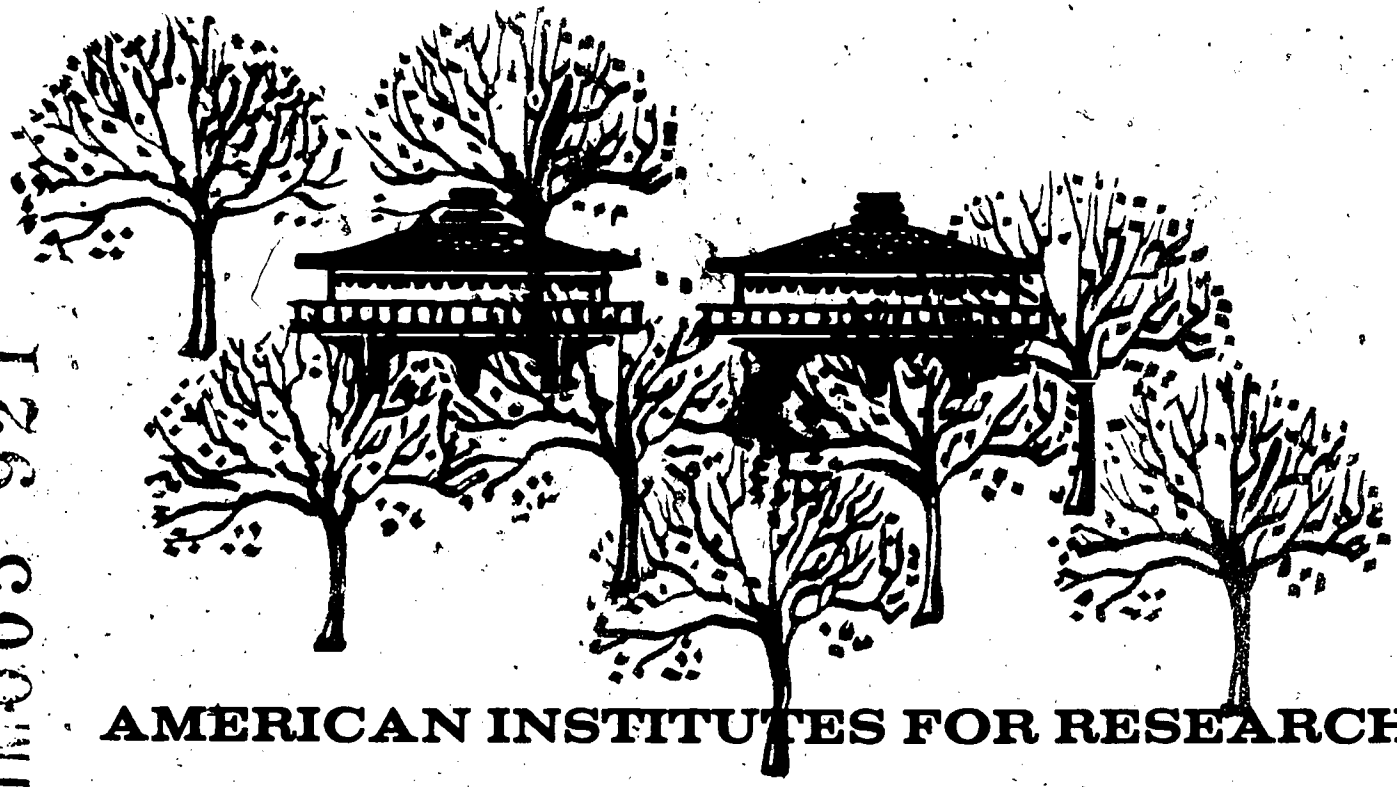
The general emphasis of Project LONGSTEP was on the identification of changes in student achievement that occur as a result of exposure to intensive educational innovation. The purpose of the analyses conducted for this supplement to Volume I of the Project LONGSTEP final report was (1) to ascertain if the trends observed and discussed in Volume I with respect to students in grades 1, 4, and 6 during the 1970-71 school year were representative of the trends shown by all analysis samples and (2) to compare results and determine if other meaningful trends across cohorts were present. The primary research hypothesis, that substantial gains in achievement test performance are positively associated with innovative emphasis, has not been supported in any general way by the analysis of Project LONGSTEP's data. These results, based on a global analysis of trends, across reading and arithmetic analysis samples and grades, tend to confirm the findings reported in Volume I for students in grades 1, 4, and 6. "Data Collection Instruments and Guidelines" developed for Project LONGSTEP referenced in Vol. I, Chapter II, Section C, will be accessioned TM 005 987 in RIEMAY77. (RC)

IMPACT OF EDUCATIONAL INNOVATION ON STUDENT PERFORMANCE:

OVERALL FINDINGS FOR READING AND ARITHMETIC

Project LONGSTEP Final Report

Volume I Supplement



AMERICAN INSTITUTES FOR RESEARCH

Palo Alto, California

AIR—21400—9/76—FR—SUPP.

IMPACT OF EDUCATIONAL INNOVATION ON STUDENT PERFORMANCE:

OVERALL FINDINGS FOR READING AND ARITHMETIC

Gary J. Coles

Albert B. Chalupsky

Project LONGSTEP Final Report: Volume I Supplement

Prepared for

Office of Planning, Budgeting and Evaluation
Office of Education
Washington, D. C. 20202

American Institutes for Research
Palo Alto, California 94302

September 1976

The research reported herein was performed pursuant to Contract No. OEC-0-70-4789 with the U. S. Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

ACKNOWLEDGMENTS

The conduct of a project as large and complex as LONGSTEP requires the assistance of a large number of agencies and individuals. The preface to Volume I of the final report attempted to acknowledge these many project contributions. The present report obviously owes its existence to all of the project contributors, and to them we again express our gratitude. In this section we would like to offer our sincere appreciation to the following individuals who made special contribution to this Volume I Supplement:

To Dr. Robert Linn, for his advice on the statistical properties of correlations involving residualized variables which served as the basis on which Attachment C was developed;

To Marion F. Shaycoft, for her technical review of the initial drafts of this report;

To Donald Reed, for his production of the figures in the report and the tables included in supporting appendices of the Volume I Supplement;

To Drs. Bruce Everett and David Gross, for their development of the analysis files and the subsequent processing of these data;

To Carolyn Davis, for supervision of the report production and most of the preparation of the manuscript;

To Edwin Carr, for his development of the summary tables contained in the Volume I Supplement and its attachments; and

To Caroline Dillman for her editorial review of the manuscript.

TABLE OF CONTENTS

List of Tables

(continued)

List of Tables (continued)

List of Figures

I. INTRODUCTION TO PROJECT LONGSTEP

Educators and noneducators alike have shown a growing awareness of the lack of--and need for--evidence as to whether or not innovative educational practices are indeed better than the more traditional approaches. In response to this need, the U. S. Office of Education in 1969 awarded a contract to the American Institutes for Research to develop a design for a study of the effectiveness of highly intensive, innovative educational practices on students in grades 1 through 12. The general emphasis of the resulting Project LONGSTEP (the Longitudinal Study of Educational Practices) was on the identification of changes in student achievement that occur as a result of intensive educational innovation, "intensive innovation" meaning the implementation of a new program encompassing a significant proportion of students, entailing a major alteration of school procedures, and involving a high investment of resources.

Specific objectives of Project LONGSTEP were to design a system to study the characteristics underlying innovative educational approaches; to establish a large-scale data base of program characteristics and student outcomes for a select sample of educational programs involving intensive and highly innovative education practices; to determine longitudinally the impact of such innovation upon student performance and attitudes; and to attempt to identify the dimensions of the components that exhibited the greatest impact on student outcomes.

A complete discussion of the project design and data collection, as well as the analytical methods and findings for three cohorts, is contained in Volume I of the final report (Coles, Chalupsky, Everett, Shaycoft, Rodabaugh, and Danoff, 1976). This supplementary report has been prepared with the expectation that the reader is familiar with Volume I.

II. OVERVIEW OF THIS REPORT SUPPLEMENT

Volume I of the final report of the Longitudinal Study of Educational Practices (Project LONGSTEP) was written to provide a detailed discussion of (1) the study's overall objectives, design, data collection procedures, instrument design and scaling, and data base and (2) the methods used and the findings obtained in an intensive analysis of the treatment and teacher correlates of reading, language and arithmetic achievement for three cohorts[1] of students--Cohorts 1, 4 and 6. This Volume I Supplement, on the other hand, describes and integrates the study's overall results obtained by means of a more global look at the reading and arithmetic achievement of all cohorts of students who participated in Project LONGSTEP. The primary objectives of this more general presentation of results are

- to discuss the findings and the conclusions based on the analyses of all cohorts and, where appropriate, compare such findings with those obtained for Cohorts 1, 4 and 6 (as described in Volume I)
- to compare the educational growth of different cohorts of students when they were at similar grade levels (during different school years) or the growth of the same group of students in two consecutive school years.

---

[1] "Cohort" is a term that is used to identify a given group of students who followed the same grade progression during the three years that the study was implemented. Cohorts are labeled by the grade level of that group of students during Year 1 of the study, the 1970-71 school year. Thus, Cohort 1 refers to all those students who were first-graders during the 1970-71 school year or who were not present in the sample during Year 1 but who would have been first-graders at that time because they were second-graders in Year 2 or third-graders in Year 3. Similarly, Cohort 4 would identify the students who were in the fourth grade in 1970-71. The term "cohort" was utilized throughout the Project LONGSTEP report to identify student groups because the study's longitudinal design meant that a given group of students would be members of three different grades, the particular grade depending on the particular school year. Table 1 reviews the grade membership of each cohort of students present in Project LONGSTEP during each of the study's three years of data collection.

TABLE 1

Grade Membership for
Students in All Cohorts by School Year.

Cohort	1970-71 (Year 1)	1971-72 (Year 2)	1972-73 (Year 3)
0	-	1	2
1	1	2	3
2	2	3	4
3	3	4	5
4	4	5	6
5	5	6	7
6	6	7	8
7	7	8	9
8	8	9	10
9	9	10	11
10	10	11	12

The findings presented in this supplement, by design, were not based on the entire sequence of analyses that were used to examine intensively the achievement of students in Cohorts 1, 4 and 6.. Rather, a subset of these approaches was used to permit a more general examination and evaluation of Project LONGSTEP's primary research hypothesis that substantial educational growth is positively associated with greater emphasis on innovation. (Due to the similarity of the Reading and Language results reported in Volume I, this supplemental report focuses only on reading and arithmetic achievement.) The remainder of this report is organized into two sections, the first of which briefly reviews the particular analytic approach upon which the findings and conclusions reported here are based. The last section summarizes the findings and contains an overview of trends across cohorts and cognitive outcomes.

III. ANALYTIC APPROACH

Analytic procedures implemented for this supplement to Volume I included most of the methods used to evaluate the impact of innovative emphasis on reading, language and arithmetic achievement in Cohorts 1, 4 and 6. These are reviewed in the remainder of this section in terms of the questions they were designed to answer.. Discussion of issues and procedures is purposely brief since most were treated in detail in Chapter IV of Volume I.

1. What general analytic approach was chosen to assess growth?

As with the Cohort 1, 4 and 6 analyses conducted for Volume I, two pretest/posttest analyses were implemented on each cohort's data. The Year 1/Year 2 analyses used the Spring 1972 test as a posttest and the Spring 1971 test as a pretest. These analyses have also been called the Spring 1972 (abbreviated as SP72) analyses throughout Volume I. The Spring 1973 (or SP73) analyses examined growth between Year 2 and Year 3, that is, between Spring 1972 (the pretest) and Spring 1973 (the posttest).

2. How was the potential analysis sample defined?

Membership in analysis samples was defined by the criteria discussed fully in Volume I. In general, the students contained in a given pretest/posttest analysis sample were those who were present in the Project LONGSTEP sample during both the pretest and posttest school years and who had no missing data with respect to the key analysis variables examined for this supplementary report. (These variables are a subset of those examined previously in Volume I and are listed in Question 4.)

3. How many students were deleted from each analysis sample because of missing data or maximum/minimum test scores?

Table 2 shows the number of students in each analysis sample and the number of students deleted from analysis because they had some form of missing data or because they obtained a near perfect score ("topouts") or near zero score ("bottomouts") on the pretest and/or posttest. Due to the complex logical and statistical problems involved in developing reasonable estimates of the impact of missing data on inferences (described more fully in Volume I), cost and time constraints did not permit special, intensive analyses to be implemented on deleted students. (Missing data and test floor/ceiling effects, however, have been discussed more fully in Volume I, Chapters IV and V.)

4. Which key variables were examined in these supplementary analyses?

Although all of the individual key analysis variables discussed in Volume I have been included in the tables and matrices of descriptive statistics accompanying this supplementary report, the findings to be discussed in Section III focus only on the relationship of intensity of innovative emphasis to growth in achievement. The measures of primary interest in the analyses reported here include

- posttest achievement, measured by the CTBS Reading Total or Arithmetic Total Expanded Scale Score corrected for site differences in time of testing (see Volume I)
- home background, measured by the Socioeconomic Status (SES) of each student's home environment

TABLE 2

Numbers of Students in Each Analysis Sample for All Cohorts

Cohort	Dependent Variable	Spring 1972 Posttest Analysis					Spring 1973 Posttest Analysis				
		Total	Complete Data	Missing Data*	Topouts	Bottom-outs	Total	Complete Data	Missing Data*	Topouts	Bottom-outs
0	Reading			NA			982	674	308	-	-
	Arithmetic						982	640	342	-	-
1	Reading	975	791	184	-	-	1049	767	282	-	-
	Arithmetic	953	752	201	-	-	1012	741	259	12	-
2	Reading	999	805	189	5	-	1057	757	278	22	-
	Arithmetic	1002	793	209	-	-	1048	699	297	52	-
3	Reading	2324	1879	400	45	-	2422	1905	476	35	6
	Arithmetic	2293	1808	368	117	-	2409	1808	601	-	-
4	Reading	2535	1952	529	54	-	2439	1925	499	15	-
	Arithmetic	2537	1943	527	67	-	2521	1964	509	48	-
5	Reading	2337	1916	396	25	-	2453	1794	567	21	71
	Arithmetic	2324	1845	423	38	18	2442	1804	554	39	45
6	Reading	2183	1520	647	16	-	1968	1552	405	11	-
	Arithmetic	2111	1443	614	54	-	1960	1505	393	62	-
7	Reading	2116	1484	613	19	-	2096	1241	720	13	122
	Arithmetic	2106	1299	635	56	116	2079	1172	738	45	124

(continued)

TABLE 2 (continued)

Cohort	Dependent Variable	Spring 1972 Posttest Analysis					Spring 1973 Posttest Analysis				
		Total	Complete Data	Missing Data	Topouts	Bottom-outs	Total	Complete Data	Missing Data	Topouts	Bottom-outs
8	Reading	2076	1406	660	10	-	1918	1238	749	6	163
	Arithmetic	2060	1295	505	47	213	1917	1069	703	30	115
9	Reading	1161	699	457	5	-	1430	846	449	14	121
	Arithmetic	1159	554	527	15	63	1420	464	850	42	64
10	Reading	1443	850	573	20	-	1290	676	492	23	99
	Arithmetic	1419	502	821	50	46	1282	215	999	48	20

Students not enrolled in reading or arithmetic classes (because they may have been electives, especially at the higher grade levels) would have had missing treatment and teacher data and therefore would have been classified as students with missing data.

- initial achievement status or pretest, measured by the appropriate subtest score obtained by each student during the previous school year (and corrected for site differences in time of testing)
- Level of Innovation, a measure of general emphasis on innovation equal to the sum of ten key indices of school practices and procedures, each index scaled so that a higher score would reflect greater judged program-level emphasis on innovation[2]
- Number of Minutes per Day, measured separately with respect to the amount of class time spent per day on language arts or mathematics activities
- Teaching Qualifications, measured by the average Teaching Qualifications scale score (an index of teacher education and experience) of a given student's teachers for each subject matter area separately (i.e., separately for each student's language arts and math teachers).

An additional outcome measure equal to the part of each student's posttest score that could not be linearly predicted from his/her pretest and SES was also computed. As noted in Volume I (Chapter IV), the analysis of the relationships between such a residualized posttest score and a set of predictor variables is based on a different educational and statistical model than an analysis of the relationships between posttest and a set of residualized predictor variables. This variable was included here so that the association between innovative emphasis and achievement growth could be examined in a slightly different way to minimize the possibility that conclusions regarding impact would be highly method dependent.

---

[2]Degree of Individualization (DI), a scale based on a subset of the items used to form Level of Innovation (LI), was not intensively analyzed as it was in Volume I because of time and cost limitations and the redundancy in the two superscales. Nevertheless, DI is included in the various correlation matrices generated for the report and reproduced in the supporting appendices.

5. How large were the pretest/posttest gains in each analysis sample and how do they compare across cohorts and grade levels?

Volume I which looked at data obtained from Cohorts 1, 4 and 6 showed that the magnitude of the yearly gains in achievement (expressed in standard deviation units--see Volume I, Equation V-1 on p. 145) decreased at each higher grade level. Although such analyses are not central to the objectives of this supplementary report, they provided a measure of average growth shown by each analysis sample and permitted a rough comparison of average gains demonstrated by different groups of students.

6. Was the posttest conspicuously farther from its norm than the pretest was from its norm?

One of the primary selection criteria for participation in Project LONGSTEP was "departure from traditional classroom practices." It was reasonable to expect, therefore, that the average gain shown by a given LONGSTEP sample should exceed that of the CTBS norm sample if a greater degree of program-level innovativeness was indeed positively related to educational growth. This question was examined by seeing if the mean posttest score of each analysis sample was notably farther from its norm (50th percentile) than the mean pretest was from its norm (50th percentile).

It should be noted that this method of examining pretest/posttest/norm differences was used to facilitate the detection of gross sample mean/norm differences across grades and years. This approach was not used to evaluate a particular sample's pretest-to-posttest growth with respect to national norms. A more appropriate test of such gain in achievement is provided by a t test (U.S. Department of Health, Education, and Welfare, Office of Education, 1976),

$$t_{N-1} = \frac{\bar{Y} - \hat{\bar{Y}}}{\sqrt{\frac{S_x^2 + S_y^2 - 2r_{xy}S_xS_y}{N-1}}} \qquad \text{(II-1)}$$

where,

$\bar{Y}$ = observed mean posttest score

$\hat{\bar{Y}}$ = expected mean posttest score (estimated from national norms on the basis of the percentile rank of the pretest mean)

S_x = pretest standard deviation

S_y = posttest standard deviation

r_{xy} = correlation between pretest and posttest scores

N = number of children

N-1 = degrees of freedom.

7. Can the different analysis samples be distinguished in terms of the Level of Innovation or Number of Minutes per Day to which they have been exposed?

If achievement growth differences among analysis samples were present to any notable degree, the treatment attributes of the groups were compared to see if the groups were, on the average, exposed to substantially different kinds of treatments in terms of emphasis on innovation.

8. What growth model was examined?

The model of achievement growth analyzed for this supplementary report was the same as that used for Cohorts 1, 4 and 6 and discussed in Volume I. Posttest achievement was hypothesized to be a function of four sets of predictors:

- student Socioeconomic Status
- initial achievement status or pretest
- Level of Innovation and Number of Minutes per Day (of class time spent on language arts or math)
- Teaching Qualifications.

Regression analysis and commonality analysis were used to examine the relationships of these predictors with each other (as joint/confounded predictors of posttest) and with posttest.

9. Were the educational experiences of students who achieved substantially more than expected for two consecutive school years notably different from those of students who achieved substantially less than expected?

The overall analyses based on all students may have shown that average group differences in achievement growth were not highly associated with intensity of innovative emphasis. It was possible, however, that some students did show dramatic gains in achievement and that they tended to be exposed to educational environments which differed systematically from the environments of students whose performance was substantially less than expected. In short, substantial associations between gains and achievement may have been masked in the overall analyses by the majority of students for whom the different treatments had no notable differential impact. For this reason, it was of interest to see if dramatic achievement gains or lack of gains on the part of individual students tended to occur in substantially different educational environments.

Students in each analysis sample whose posttest residual score (residualized on the basis of pretest and SES) was equal to or greater than one-half of a standard deviation from the mean of the residuals were identified. The SP72 and SP73 samples for each cohort were merged, and the students who achieved more than expected during both school years (i.e., the "positive outliers" or "overachievers") and the students who achieved less than expected (i.e., the "negative outliers" or "underachievers") were selected for analysis. A dummy variable encoding group membership was then assigned to each student (overachievement = 2; underachievement = 1), and this dummy variable was correlated with the various key analysis variables. The resulting point-biserial correlations provided an index of the mean difference between the groups on each key analysis variable. The correlations were then examined to see if overachievement or underachievement in two consecutive school years was highly associated with attributes of the educational treatments to which such students had been exposed. As with similar analyses described in Volume I, significance tests were not performed because of the highly select nature of subsamples analyzed.

10. What approach was used to examine overall trends in the data?

To facilitate examining the data for overall trends, the results based upon a given statistical technique were summarized in cohort-by-grade matrix form like the table of posttest means shown in Table 3. (The statistical analyses summarized in these cohort-by-grade tables have not been included in the body of the report--see Attachment A for the location of these complete tables.) Examination of Table 3 shows the following:

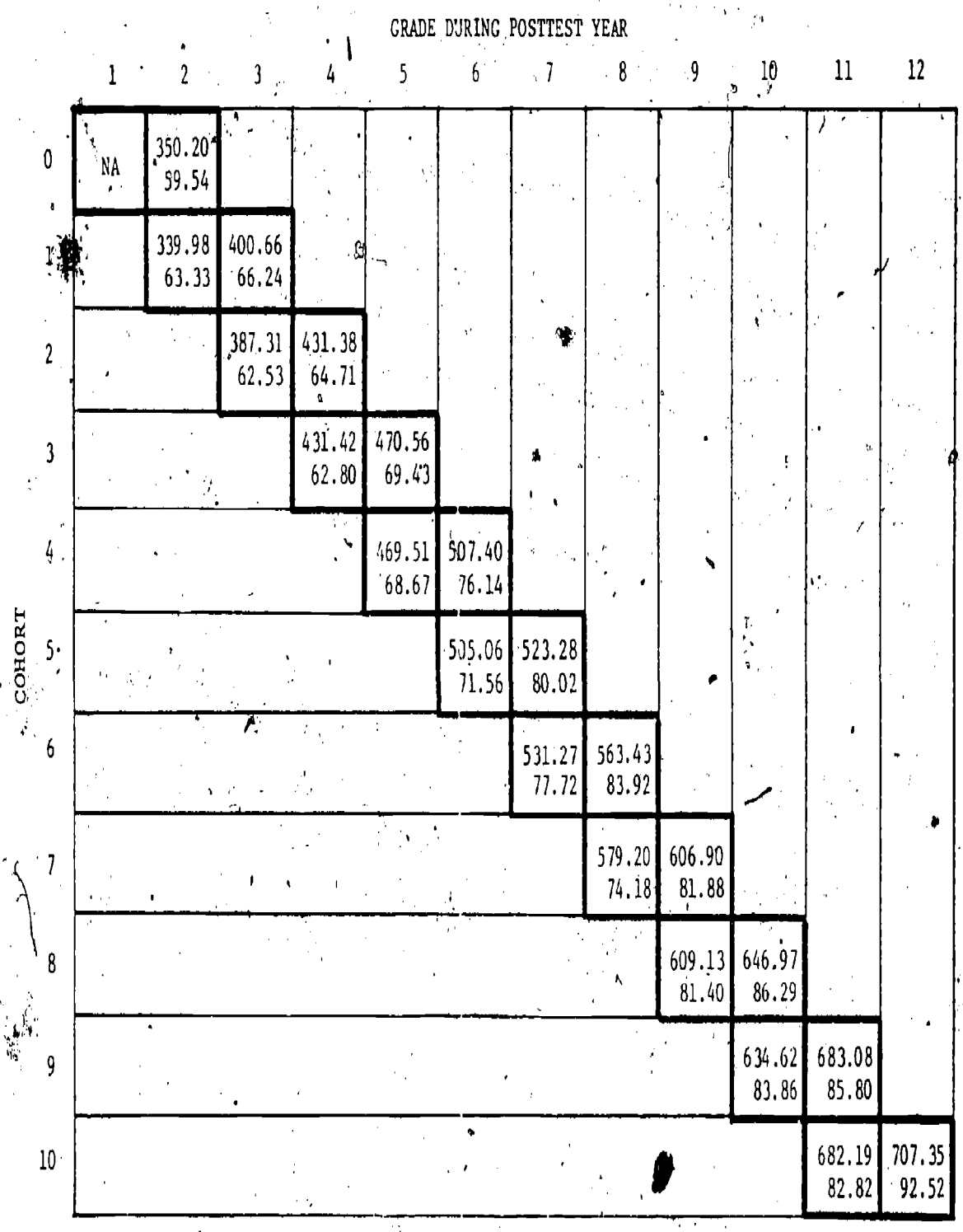
- The results or descriptive statistics present in two adjacent cells in the same row pertain to the performance of the same cohort of students, but during two consecutive school years, 1971-72 and 1972-73[3].
- The results or descriptive statistics present in adjacent cells in the same column pertain to the performance of different cohorts of students during the same school year.
- Only the two cells along the major diagonal are filled because Project LONGSTEP collected posttest data during only two school years, 1971-72 and 1972-73.

This particular manner of summarizing the findings was utilized because it enabled us to examine and compare the performance of a given cohort across school years and grade levels and to relate this performance to changes in school environments in different school years. Such an approach may be viewed as an approximation to a "within-subjects" longitudinal analysis since the student composition of the two analysis samples (1971-72, 1972-73) for a given cohort was fairly similar (see footnote below).

[3] To make maximum use of the data available, all students present during Year 2 (1971-72) or during Year 3 (1972-73) who had a valid pretest score from the previous school year and no missing data on the key analysis variable were included in a given analysis sample. Therefore, the student composition of these samples could vary somewhat.

TABLE 3

Posttest Means and Standard Deviations - Reading Total Score Analyses



Furthermore, this mode of reviewing the study's results also permits a "between-subjects" form of analysis in which the performance of different student groups at the same grade (but during different school years) can be compared and related to changes in school environments. Since adjacent cohorts of students attended approximately the same schools, the approach used should be helpful in determining whether or not treatment impact is unique to a given group of students and/or school year.

IV. RESULTS AND DISCUSSION

Review of Objectives and Approach

The overall objective of this supplement to Volume I of the Project LONGSTEP final report has been to assess trends in achievement growth, across grades, across groups/cohorts of students, and across the study's two primary cognitive outcomes--reading and arithmetic achievement. The specific analytic objectives developed to achieve this overall goal were

- to determine, across all grades and school years, if the average posttest reading and mathematics performance present in LONGSTEP's sample of fairly innovative schools was conspicuously farther from national norms than was the samples' average pretest scores
- to determine, across all grades and school years, if variation among and within analysis samples with respect to growth in reading and mathematics achievement was positively associated with variation in program-level emphasis on innovation
- to determine if those individual students who achieved more than expected (on the basis of their SES and previous year's achievement level) during two consecutive school years were exposed to programs with greater emphasis on innovation than were students who achieved less than expected.

Findings for Reading Achievement

Average achievement differences among analysis samples. Results of the reading achievement analyses showed that the average posttest performance (see Table 3) at each grade level was fairly similar, even when different groups (i.e., cohorts) of students were involved during consecutive school years. For example, the mean posttest Reading Total Score (Table 3) for different groups of sixth-graders was 507 (for students in Cohort 4 during 1972-73) and 505 (for students in Cohort 5 during the 1971-72 school year). As shown in Figure 1, with the possible exception of growth during the third grade (i.e., between grades 2.7 and 3.7) and during the eighth grade, average posttest scores for the Project LONGSTEP analysis samples were not farther from national norms than were their average pretest scores. Although average pretest and posttest reading achievement for all analysis samples were above national norms (except for Cohort 5 during the seventh grade), the reading gains shown by the set of fairly innovative schools participating in the study did not notably exceed the growth expected on the basis of CTBS norms.[4]

Table 4 contains the pretest-to-posttest gains for each analysis sample. Inspection of these results suggests that the largest gains in reading seemed to have occurred in the elementary grades. The following "irregularities" also seem to be apparent.

- The average gain shown by students in Cohort 5 during the 1971-72 school year (when they were sixth-graders) was notably larger than their gain the following school year (when they were seventh-graders).
- Cohorts 6 and 7 demonstrated different amounts of growth (on the average) during the eighth grade, but Cohort 7's average gain decreased somewhat during grade 9.

[4]The CTBS pretest and posttest national norms for Reading Total Scores and the number of students in each analysis sample are provided in Attachment B, Tables B-1, B-2 and B-3, respectively.

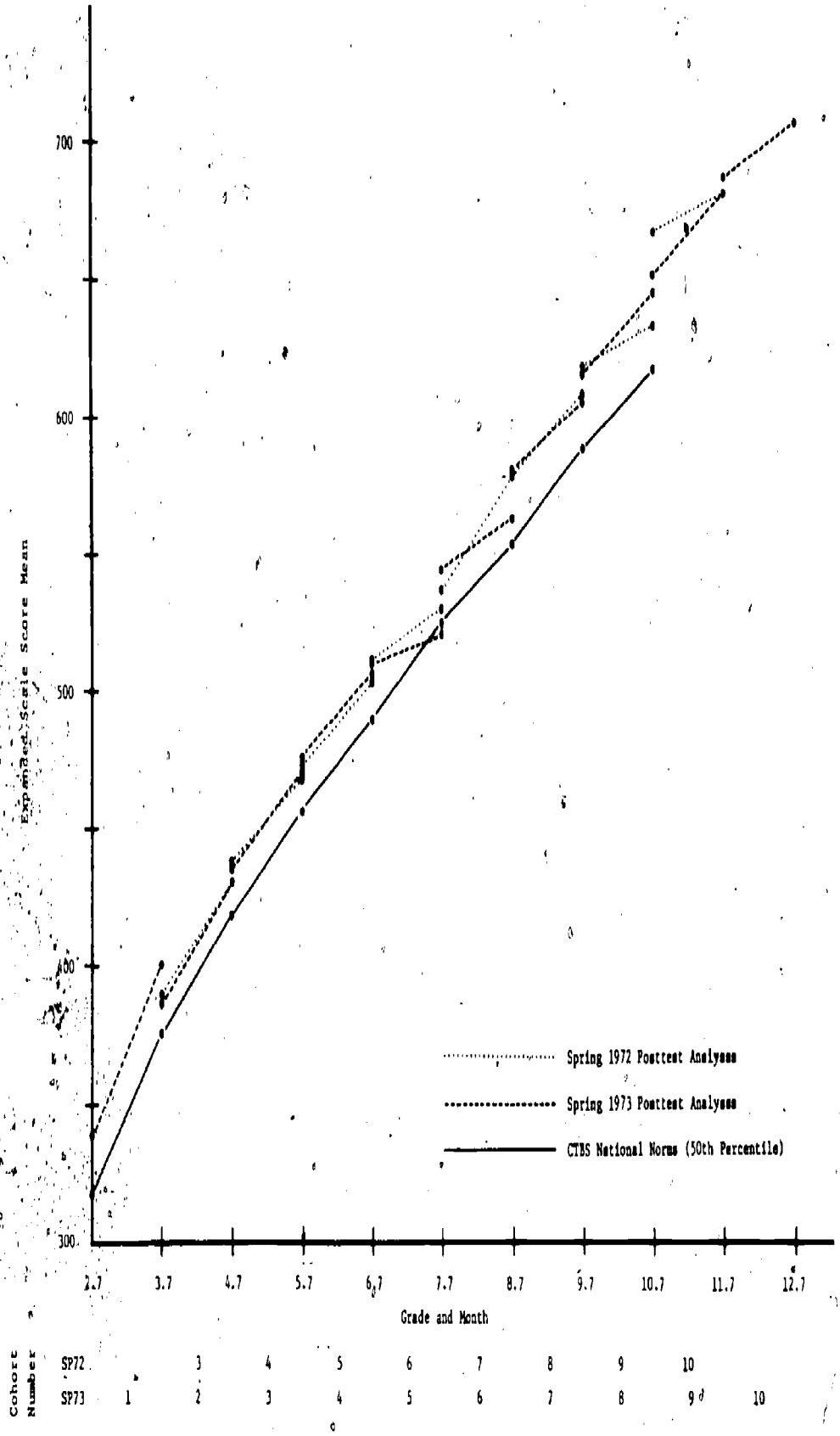


Figure 1. Pretest and posttest means and published national norms (50th percentile for spring testing) - Reading Total Score analyses.

TABLE 4

Mean Differences Between Reading Total Pretest and Posttest Scores and "Standardized Gains"[1]

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA[2]	NA[2]										
1		NA[2]	62.20 .96									
2			NA[2]	45.27 .71								
3				41.38 .65	35.49 .54							
4					31.60 .48	31.67 .44						
5						31.24 .44	10.97 .15					
6							17.97 .24	17.98 .23				
7								41.20 .56	26.07 .34			
8									29.88 .39	30.09 .37		
9										14.78 .18	30.70 .39	
10											14.13 .17	19.23 .23

[1]"Standardized gains" are pretest/posttest mean differences expressed in average or pooled standard deviation units--see Volume I, Equation V-1, on page 145.

[2]Not computed because the CTMM was used as the pretest.

- Average gains during grade 10 were notably larger in Cohort 8 than in Cohort 9, but Cohort 9's average gain increased somewhat during grade 11.

Many of the same results were also evident when average pretest and post-test reading achievement were compared with CTBS national norms (50th percentiles). Figure 1 shows these trends across analysis samples.

Six possible reasons for such differences among group means were briefly considered--group/cohort differences with respect to

- average pretest level
- average student Socioeconomic Status
- average Level of Innovation
- average Number of Minutes per Day (of class time on language arts activities)
- average Teaching Qualifications
- schools attended.

Examination of Table B-4 in Attachment B shows that there were no large differences between cohorts with respect to average pretest level at the same grade level. Neither were there large group differences with respect to average SES (see Attachment B, Table B-5). Although there were differences with respect to mean Level of Innovation (Attachment B, Table B-6) and Teaching Qualifications (Attachment B, Table B-7) to which the various analysis samples were exposed, brief examination did not suggest that any consistent relationship existed between these mean attributes and the irregularities in average gains noted previously.

Except for the lack of hypothesized, conspicuous positive growth in achievement, the most systematic relationship across the various Project LONGSTEP analysis samples seems to be between average Number of Minutes per Day (of class time on language arts activities) and average achievement gains. Table 5 shows that students in grades 1 through 6 were exposed to about one and a half hours per day (on the average) of class activities devoted to language arts. Beginning with the seventh grade, however, students who were enrolled in "language arts" kinds of classes

TABLE 5

Number of Minutes per Day Means and Standard Deviations - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	101.05 32.15										
1		96.83 27.76	95.66 34.66									
2			103.55 32.41	95.81 34.40								
3				100.44 29.69	100.35 26.51							
4					94.93 26.75	81.93 26.34						
5						83.80 27.01	53.80 10.05					
6							55.22 10.86	50.51 8.22				
7								55.42 10.96	54.05 5.83			
8									54.43 5.78	55.07 7.01		
9										53.84 8.59	54.27 7.18	
10											57.20 20.25	48.61 15.55

were exposed to less than one hour of language arts per day--about one 45-minute to 60-minute class period per day. Since there also seems to be a drop in mean achievement gains between the sixth and seventh grades, it is tempting to speculate that less exposure to instruction (at least as measured by the index Number of Minutes per Day) at the junior high and senior high school levels may be one of the reasons the gains demonstrated by these Project LONGSTEP grades were less than the average gains shown by our elementary grades. However, a number of possible alternative explanations for this trend should be considered.

Perhaps one of the most immediate explanations for the average reading achievement gains to be less at the junior high school and senior high school levels is that the students who were enrolled in language arts classes in these grades may have tended to be in remedial classes, especially in the last two years of high school. This is not a completely reasonable explanation for the trends observed here, however, because all Project LONGSTEP analysis sample pretest means were above national norms (see Figure 1). Since the students included in each analysis sample were only those who were receiving language arts instruction (i.e., they had no missing data on the Level of Innovation index or the Teaching Qualifications of their language arts teachers), it is more correct to assert that all LONGSTEP analysis samples, on the average, were comprised of above average readers (relative to the 50th percentile norm). In fact, Figure 1 even suggests that students in the junior and senior high school grades, on the average, had slightly higher pretest scores (relative to norms) than did students in the elementary grades.

A second possible alternative explanation for reading gains to be less in the higher grades is that the elementary schools participating in Project LONGSTEP were notably more effective than were the junior/senior high schools. As explained more fully in Volume I, the particular grades selected and tested at a given site/school depended on the specific grades at which the site's special program(s) were targeted. Table B-8 in Attachment B shows the school location for all cohorts of students during Year 2 and Year 3 of the study. Examination of this table shows that

- the schools in which members of the same cohort could have been enrolled were not always the same during the 1971-72 and 1972-73 school years
- the school membership of the cohorts differed somewhat, especially in consecutive school years.

Thus, some of the small fluctuations in growth in average reading achievement among analysis samples could have been due to the fact that the students in two analysis samples (either the same cohort during two consecutive school years and at two different grade levels or different cohorts during two consecutive school years but at the same grade level) were members of a somewhat different set of schools. Adjacent cohorts, however, tended to be members of the same or almost the same schools.

Another possible explanation is that language arts activities in the elementary grades may be qualitatively different from such activities in junior/senior high school. "Language arts" during the elementary grade years undoubtedly involves primary emphasis on reading skills, whereas "language arts" in junior/senior high schools usually involves a much broader set of elective courses or "enrichment" activities. Thus, instructional or class time differences between the high school and elementary grades could be confounded with a curriculum content dimension, and it is this factor that may be the crucial variable.

Finally, it is also possible that the overall trend for reading gains to decrease during the later grades may be due, in part, to (1) the particular manner in which the CTBS Reading Total Scores were scaled (Expanded Scale Scores provided by the test publisher were used) and/or (2) the tendency for the acquisition of reading skills to follow the classical learning curve. Since exposure to language arts activities also decreased in the later grades, it is possible that the observed across-analysis-sample association between growth in reading achievement and Number of Minutes per Day may be merely a result of concomitant variation.

In spite of the other possible explanations for the concomitant decrease in average reading and in exposure to language arts activities, it is intriguing to note the presence of such an across-

analysis-sample trend in the Project LONGSTEP data. The fact that this loss in rate of cognitive growth was most dramatically demonstrated by a fairly constant group of students (Cohort 5) during the years they moved from elementary to junior high school certainly suggests that changes in reading curricula and school membership, as well as in instructional time, may be important factors to consider in optimizing reading achievement growth in the higher grades. The following section examines the relationship of Level of Innovation, Number of Minutes per Day and Teaching Qualifications to achievement within each analysis sample.

Associations with achievement within analysis samples. Commonality analyses (see Attachment B, Tables B-9, B-10, B-11 and B-12) of the four sets of predictors in the growth model (pretest, SES, Level of Innovation and Number of Minutes per Day, and Teaching Qualifications) showed that very little posttest variance was uniquely associated with either of the two sets of treatment measures--Level of Innovation and Number of Minutes per Day (as a set) and Teaching Qualifications. In fact, in only one analysis sample (Cohort 1, SP73) was the uniqueness for either set of treatment variables larger than .02. In this one sample, on the other hand, almost nine percent of the variance in reading achievement posttest scores for third graders (for Cohort 1 during the 1972-73 school year) could be attributed uniquely to Level of Innovation and Number of Minutes per Day. Nevertheless, the uniquenesses show that there was no general trend within analysis samples (i.e., involving different grades and cohorts/groups of students) for there to be substantial linear associations between growth in reading achievement and unique variation in Project LONGSTEP's primary treatment indices, i.e., program-level emphasis on innovation (Level of Innovation) and typical class time spent on language arts activities (Number of Minutes per Day).

Examination of the growth model regression analyses also showed that neither Level of Innovation nor Number of Minutes per Day tended to receive statistically significant coefficients that were of the same sign across analysis samples (i.e., across different grades and cohorts). These results are summarized in Tables 6 and 7. Therefore, program-level

TABLE 6

Sign and Statistical Significance of the Level of Innovation Regression Coefficients - Reading Total Score Analyses

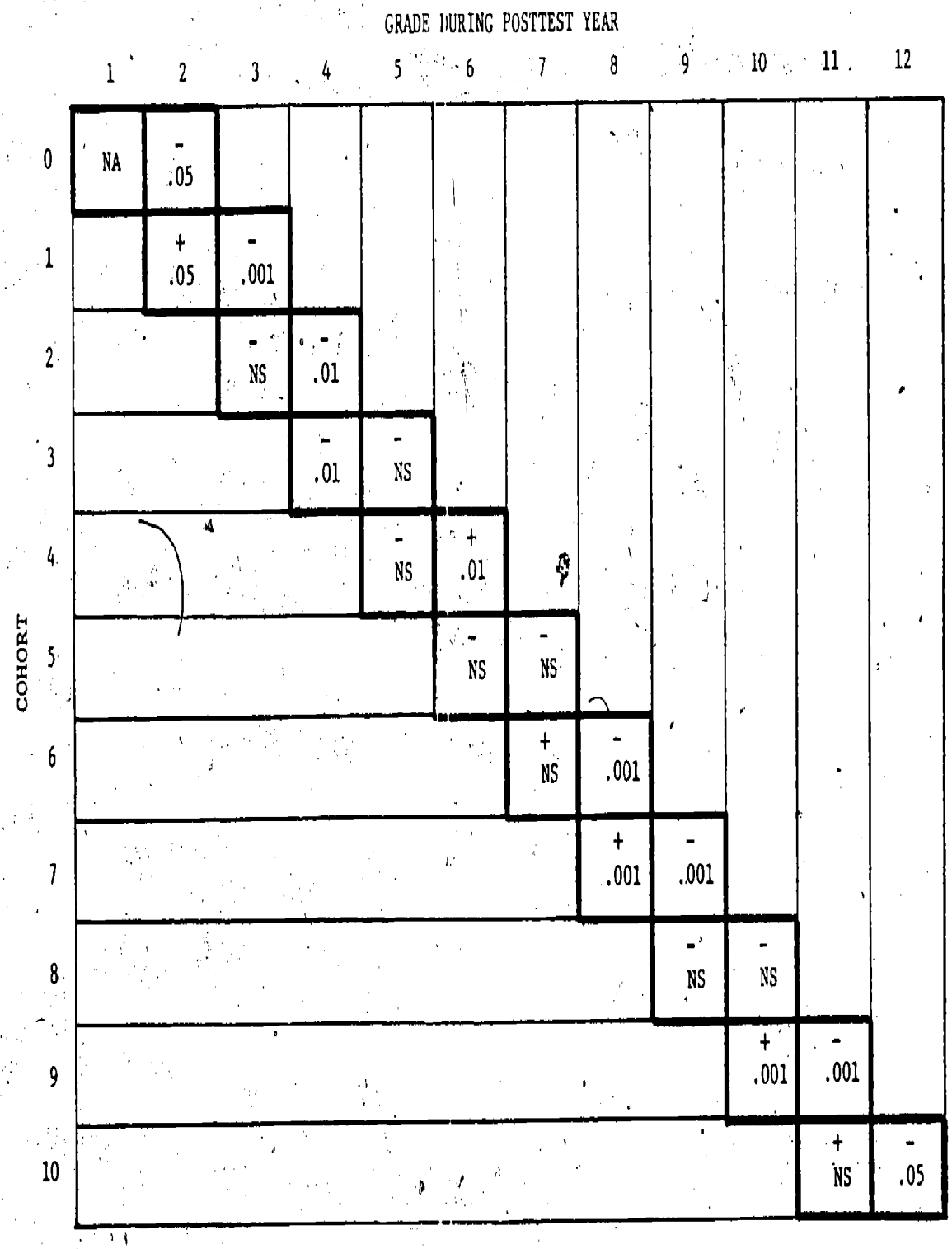
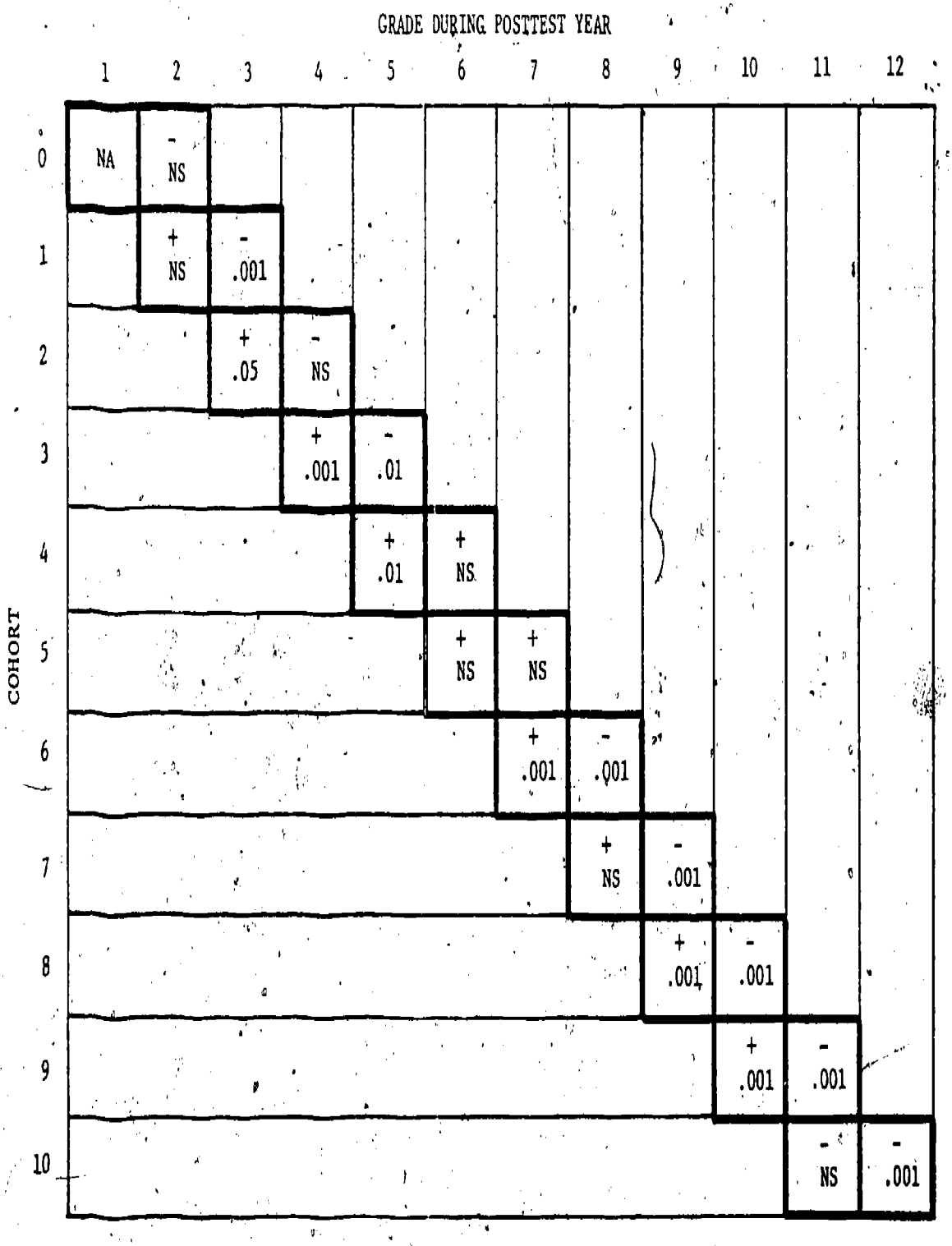


TABLE 7

Sign and Statistical Significance of the Number of Minutes per Day Regression Coefficients - Reading Total Score Analyses



emphasis on innovation neither accounted for substantial amounts of post-test variance within this set of analysis samples nor did it have small associations that were consistently positive across different analysis data sets (i.e., across grades and groups/cohorts of students).

Table B-12 in Attachment B shows that the unique component of Project LONGSTEP's teacher index, Teaching Qualifications, was not highly associated with posttest reading achievement; that is, it accounted for almost no posttest variation independently of pretest, SES, Level of Innovation and Number of Minutes per Day. Table 8 shows that the regression coefficients for Teaching Qualifications in our growth model also showed no obvious tendency to be statistical significant and of the same sign across analyses.

Lack of consistent relationships across grades and cohorts/groups of students was also shown in the correlations of Level of Innovation, Number of Minutes per Day and Teaching Qualifications with a posttest score residualized on the basis of pretest and SES (technically, a residualized growth/gain score). Tables 9, 10 and 11 summarize these correlations. The within-analysis-sample trends, then, also have not shown that program-level emphasis on innovation was highly or consistently related to growth in reading achievement in any general manner.

As noted in the discussion of methods in Section II and in Volume I, it is possible that program-level emphasis on innovation may not be highly associated with achievement since it may not have large average effects--that is, it may not be substantially related to achievement for the majority of students in a particular sample. It is certainly possible, however, that the more innovative approaches sampled by Project LONGSTEP may have been extremely important for some students. To detect such associations, those students who achieved more than expected on the basis of their particular pretest and SES scores (during two consecutive school years) were identified. A comparison group was formed from those students who showed notable lack of growth during two consecutive school years. The numbers of students identified and compared were not large (see Table B-13 in Attachment B). Nevertheless, the "overachievers" and "underachievers" differed dramatically with respect to average reading achievement during

TABLE 8

Sign and Statistical Significance of the Teaching Qualifications Regression Coefficients - Reading Total Score Analyses

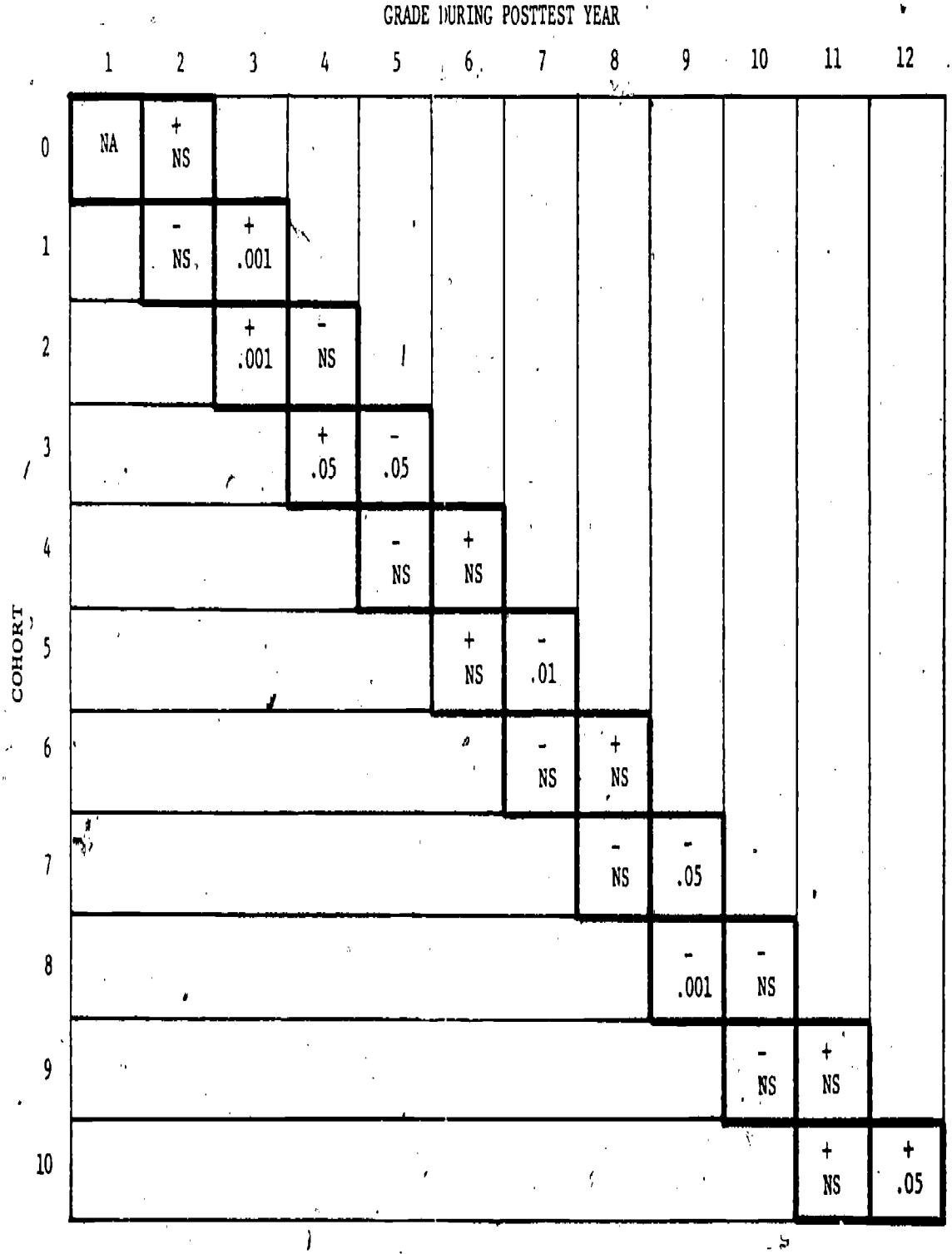


TABLE 9

Correlations Between a Residualized Posttest Score and Level of Innovation - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR											
	1	2	3	4	5	6	7	8	9	10	11	12
0	NA	-.079										
1		.069	-.352									
2			-.021	-.060								
3				-.032	-.013							
4					-.027	.063						
5						-.023	.008					
6							.057	-.139				
7								.113	-.120			
8									-.004	.160		
9										.097	.018	
10											.066	.122

TABLE 10

Correlations Between a Residualized Posttest Score and Number of Minutes per Day - Reading Total Score Analyses

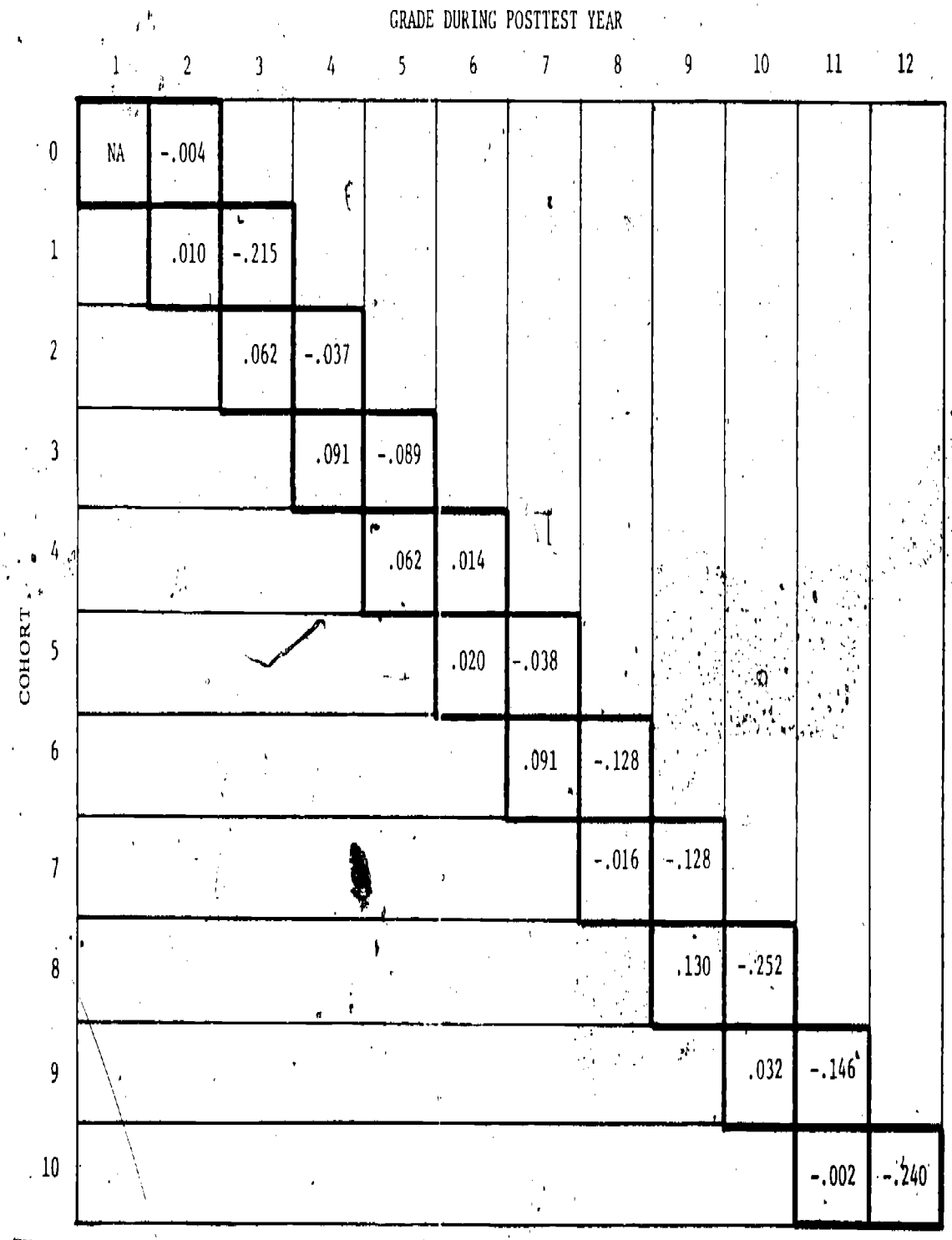
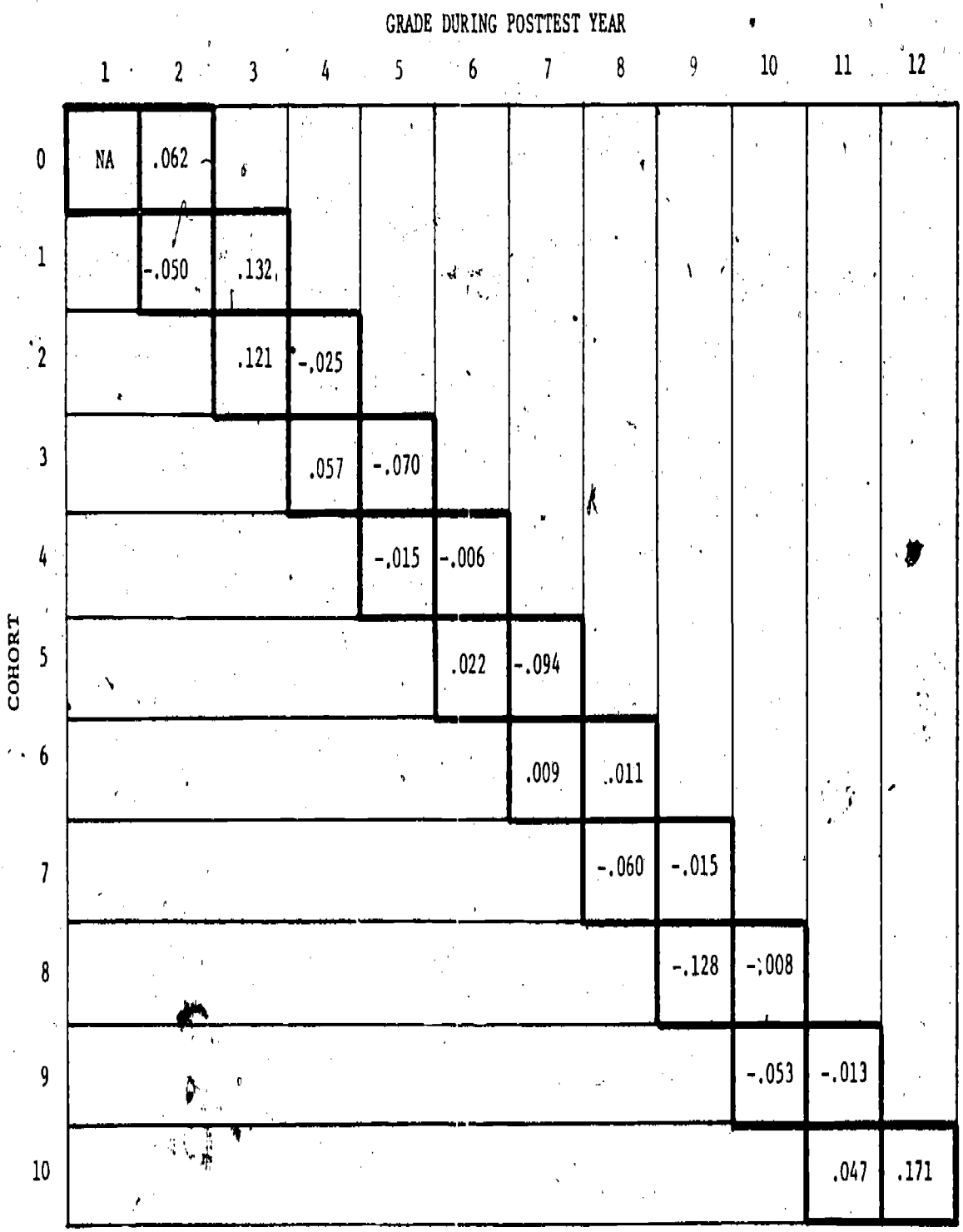


TABLE 11

Correlations Between a Residualized Posttest Score and Teaching Qualifications - Reading Total Score Analyses



Year 3 of Project LONGSTEP and certainly represent students who, for two years in a row, did either much better or much worse than their achievement level or home background would predict. A point-biserial correlation between overachiever/underachiever group membership and Year 1 test performance was computed to provide a measure of the relative difference between the means for these two groups. Table B-14 in Attachment B shows that the overachievers did, on the average, tend to have somewhat higher Year 1 test scores than did the underachievers. SES differences among these two groups of students were not so notable. Table B-15 in Attachment B shows that in only two cohorts/groups of students was there an SES difference among achievement groups worthy of note (Cohorts 3 and 5). In these two groups, overachievement was associated with higher SES levels (relative to these overachieving and underachieving students, at least).

Mean differences in the Year 2 and especially the Year 3 test scores of the overachievers and underachievers, however, were dramatic (see Table B-16 in Attachment B). The magnitude of difference, as indexed by the point-biserial correlations, seems to suggest that consistent overachievement and underachievement, at least as defined here, is not highly associated with initial achievement status or SES across cohorts.

Point-biserial correlations between overachiever/underachiever group membership and the treatment variables and teacher index in the growth model show that underachievement or overachievement was not highly and consistently related to either Level of Innovation, Number of Minutes per Day, or Teaching Qualifications across all analysis samples (Attachment B, Tables B-17, B-18 and B-19). Although there did not appear to be any dramatic overall trend across all cohorts, the results did seem to suggest the following, relative to underachievers:

- Overachievers who were in elementary and junior high school[illegible] in Year 1 (Cohorts 1 to 8) were members of programs that had a lower Level of Innovation (on the average) during Year 2 and again during Year 3.

- Overachievers who were in high school during Year 1 (Cohorts 9 and 10) were members of programs that had a higher Level of Innovation (on the average) during Year 2 and again during Year 3.
- Overachievers who were in grades 1 to 5 during Year 1 were exposed to more class time on language arts activities than were underachievers during Year 2 but less time (relative to underachievers) in Year 3.
- Overachievers in the older cohorts were exposed to less class time on language arts activities during Year 2.
- Overachievers in all cohorts were exposed to less class time on language arts during Year 3.
- Overachievers in many of the cohorts were exposed to a teacher (or teachers) with a higher Teaching Qualifications score during at least one of the two posttest school years (i.e., either Year 2 or Year 3).

Although the within-analysis-sample findings reviewed in this section do not tend to support the hypothesis that greater program-level emphasis on innovation will, in general, lead to substantial improvements in reading achievement, there were two findings which seemed particularly worthy of note:

- the large unique impact of Level of Innovation and Number of Minutes per Day for Cohort 1 students during the third grade
- the slight tendency across procedures for growth in reading achievement during Year 2 to be positively related to the three treatment/teacher indices in the growth model (especially Number of Minutes per Day) and for Year 3 growth to be negatively related to the same indices.

Our presentation of Project LONGSTEP's findings for reading achievement will close with a brief discussion of these two points.

Table B-11 in Attachment B shows that the uniqueness for Level of Innovation and Number of Minutes per Day (as a set) accounted for about nine percent of the variance in the CTBS Reading Total Score posttest for Cohort 1 students when they were in the third grade (during the 1972-73 school year). Both predictors also received negative regression coefficients (see Tables 6 and 7) in the growth model regression analysis of this cohort's data. These same findings, however, were not observed during the previous school year, 1971-72, for Cohort 1 students during the second grade or for Cohort 2 students during the third grade. Since students in Cohort 1 and Cohort 2 attended the same schools as second- and third-graders, it seems possible that the combination of Level of Innovation and Number of Minutes per Day was confounded with some set of educational practices or school personnel factors to which the third-graders in some schools were exposed during the 1972-73 school year and which had a notable impact on achievement. Unfortunately, the analyses utilized for this report were not designed to "tease out" such effects.

The apparent trend for growth in reading achievement to be positively associated with Number of Minutes per Day in the 1971-72 posttest analyses and negatively associated with the same index in the 1972-73 posttest analyses was considered in light of the fact that the Year 2 test was both a posttest (in the Year 2 posttest analyses) and a pretest (in the Year 3 posttest analyses). It can be shown that under certain circumstances, the correlation of a variable like Number of Minutes per Day (NMIN) with a Year 2 posttest residual score will be positive while the correlation of that same variable with a Year 3 posttest residual score will be negative (see Attachment C for a short technical discussion). In general, this pattern of results could be expected if the slope of the regression of Year 3 test scores on Year 2 test scores was near 1.0 and the Year 2 test/NMIN covariance was larger than the covariance of NMIN with the Year 1 or Year 2 tests.

The correlations with Number of Minutes per Day in the Year 2 and Year 3 analysis samples for a given cohort, however, show that the Year 2 test/NMIN correlations in the Year 2 analyses (where the Year 2 test was

the posttest) were not always similar to the Year 2 test/NMIN correlations in the Year 3 analyses (where the Year 2 test was the pretest). Thus, it seems appropriate to assert that the opposite impact that Number of Minutes per Day had in consecutive school years probably was not an artifact completely due to use of the Year 2 test as both a posttest and as a pretest. A more appropriate and reasonable speculation is that exposure to language arts instruction varied across school years, along with changes with respect to the specific students exposed to different amounts of instruction in consecutive school years and with respect to other attributes of the program of which students were members.

Findings for Arithmetic Achievement

Average achievement differences among analysis samples. Average posttest arithmetic performance was fairly similar for different cohorts/groups of students at the same grade level (see Table 12) during consecutive school years. The only exceptions worthy of note were the mean differences between Cohorts 5 and 6 during the seventh grade and between Cohort 6 and 7 during the eighth grade. Pretest means (see Attachment D, Table D-4) appeared to be even more similar across cohorts than were the posttest means. Nevertheless, with the possible exception of the growth demonstrated by Cohort 1, students during the third grade (in the 1972-73 school ar), average posttest CTBS Arithmetic Total Scores were not conspicuously farther from national norms (50th percentile) than were their average pretest scores. CTBS pretest and posttest means are displayed for all samples in Figure 2.[5] Since one of the major criteria for selection to participate in Project LONGSTEP was departure from "traditional" classroom practices, these overall results do not support the hypothesis that greater program-level emphasis on innovation is positively related to notable gains in arithmetic achievement.

Figure 2, however, does show that the average gains in arithmetic achievement during a given grade level were not always identical for different cohorts of students. Gains during consecutive school years also

[5]The CTBS pretest and posttest national norms for Arithmetic Total Scores and the number of students in each analysis sample are provided in Attachment D, Tables D-1, D-2 and D-3, respectively.

TABLE 12

Posttest Means and Standard Deviations - Arithmetic Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	327.90 36.07										
1		320.02 38.90	377.79 42.41									
2			371.47 44.37	402.94 41.08								
3				405.19 38.94	450.27 56.94							
4					446.66 56.43	480.84 61.58						
5						479.57 61.56	496.13 73.79					
6							510.30 72.52	537.81 73.94				
7								552.98 72.18	595.68 79.49			
8									598.51 75.41	627.51 78.56		
9										627.49 81.02	659.84 74.29	
10											668.66 77.46	683.92 66.89

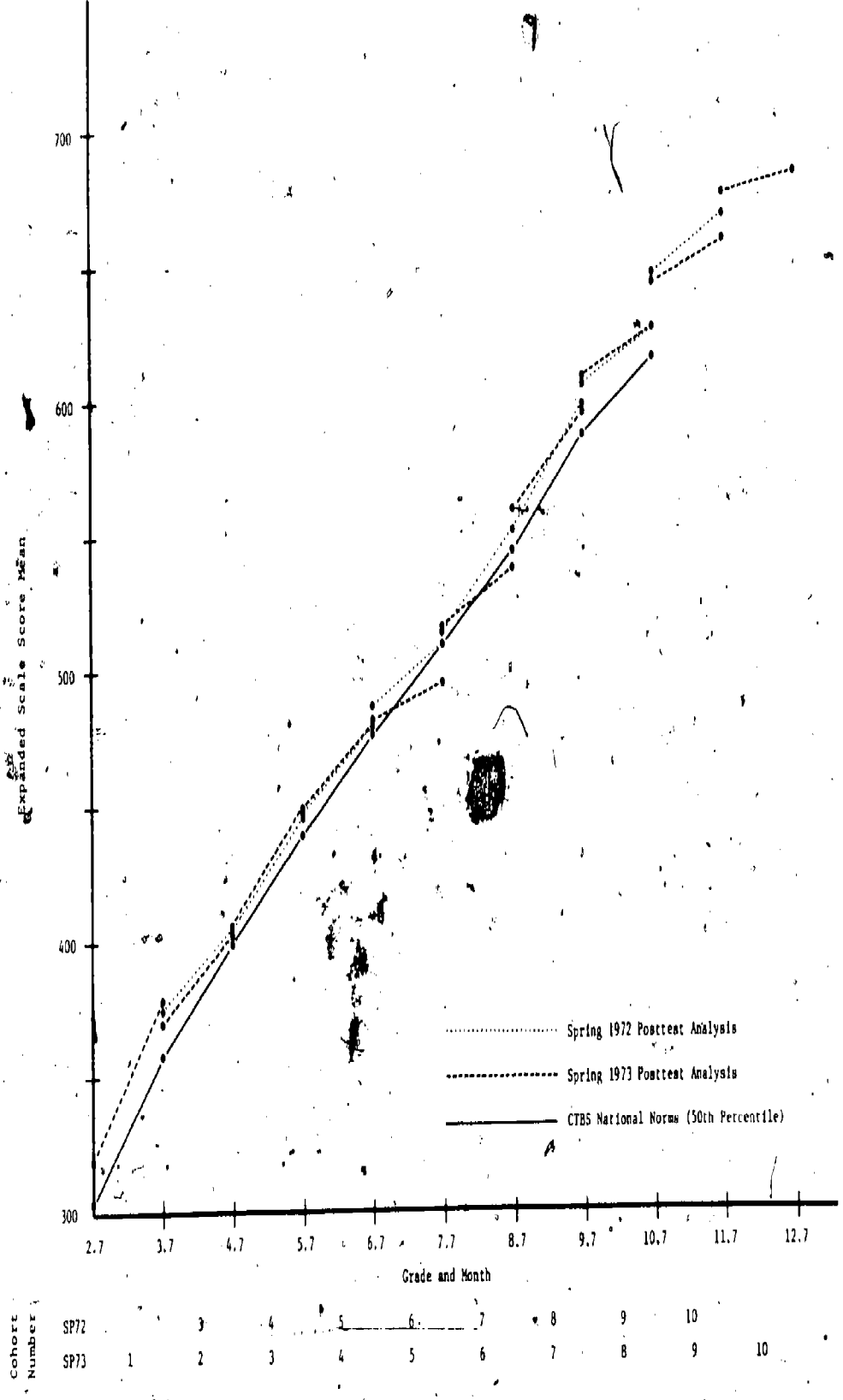


Figure 2. Pretest and posttest means and published national norms (50th percentile for spring testing) - Arithmetic Total Score analyses.

were not necessarily identical for the same cohort of students. Across-analysis-sample differences were observed both with respect to average gains (see Table 13) and with respect to pretest/posttest/norm differences.

As noted previously, there were no large differences among analysis samples with respect to average pretest level (see Attachment D, Table D-4). Thus, the small fluctuations in mean arithmetic posttest that were observed reflected different mean gains in a number of analysis samples. Inspection of the pretest-to-posttest growth for the various analysis samples (see Table 13) shows that the largest mean gains (relative to their standard deviations) seem to be occurring in the earlier grades.

In addition, Table 13 also shows the following irregular differences among analysis samples:

- The average arithmetic gain shown by students in Cohort 5 during the 1971-72 school year (when they were sixth-graders) was notably larger than their average gain the following school year (when they were seventh-graders); however, Cohort 6 students did not demonstrate the same achievement losses during the seventh grade.
- Cohorts 6 and 7 showed somewhat different amounts of average growth during the eighth grade.
- The average gain in arithmetic performance for students in Cohort 8 were larger when they were ninth-graders than when they were tenth-graders.

An examination of Figure 2 also suggested the presence of greater pretest-to-posttest growth during grade 5 in relation to grade 4 and grade 6 growth. Project staff examined the analysis samples showing such differences to see if differences among samples were related to similar variation with respect to

- average pretest level
- average student Socioeconomic Status
- average Level of Innovation

TABLE 13

Mean Differences Between Arithmetic Total Pretest and Posttest Scores and "Standardized Gains"[1]

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	NA										
1	NA	NA	58.64 1.47									
2			NA	33.25 .79								
3				30.59 .80	43.35 .89							
4					41.97 .87	31.40 .53						
5						30.94 .52	13.67 .21					
6							22.84 .35	20.64 .29				
7								38.41 .55	35.06 .48			
8									45.57 .67	18.13 .24		
9										20.87 .27	16.30 .23	
10											21.18 .29	6.98 .10

[1] See Volume I, Equation V-1, p. 145.

- average Number of Minutes per Day (of class time devoted to arithmetic activities)
- average Teaching Qualifications
- schools attended.

Mean differences among analysis samples with respect to SES, Level of Innovation, Number of Minutes per Day, and Teaching Qualifications (see Attachment D, Tables D-5, D-6, D-7 and D-8) did not seem to show any consistent relationship with the across-analysis-sample irregularities or trends noted in Table 13 with respect to the average arithmetic achievement of each analysis sample. For example, the decrease of average growth for Cohort 5 students during the seventh grade relative to average growth during the sixth grade was not associated with notable mean differences in mean SES, Level of Innovation, Number of Minutes per Day, or Teaching Qualifications. The most likely explanation for the decline for these students is that, as expected, the students in Cohort 5 attended as seventh-graders schools different from those they attended as sixth-graders (see Attachment B, Table B-8). It seems conceivable that changes not measured by the primary variables of interest in this report probably accounted for the differences observed. Similarly, the fact that Cohort 6 students as seventh-graders the year before did not show the same decline is probably due to the different school composition of the Cohort 5 and Cohort 6 seventh-grade analysis samples. As shown in Attachment D, Table D-6, however, Cohort 6 seventh-grade students were members of EdExAG groups which, on the average, had a greater program-level emphasis on innovation than did the EdExAG groups for Cohort 5 seventh-graders.

The differences in mean arithmetic achievement gains between Cohort 6 and Cohort 7 students during the eighth grade, on the other hand, showed exactly the opposite association with innovative emphasis. Although students in both cohorts were members of the same schools as eighth-graders (see Attachment B, Table B-8), greater average gains (during 1971-72 when the Cohort 7 students were eighth-graders) were associated with a lower program-level emphasis on innovation (on the average). Conversely, the difference in average yearly achievement gains during grade 9 and grade 10

for students in Cohort 8 was also associated with only a small mean difference with respect to Level of Innovation and Teaching Qualifications. Greater mean gains for those students was accompanied by a slightly higher mean Level of Innovation and Teaching Qualifications. Movement of students in Cohort 8 from junior high to senior high schools between the ninth and tenth grades in some sites could also have affected the results (see Attachment B, Table B-8). Interestingly enough, Cohort 9 students as tenth-graders in 1971-72 were exposed to much more emphasis on innovation (on the average) than were Cohort 8 students who were in the tenth grade the next year (1972-73)--yet, the gains for the two cohorts during tenth grade were very similar (see Figure 2).

Thus, the across-analysis-sample mean differences that were present in Project LONGSTEP also do not support the speculation that small, but consistent, arithmetic achievement differences among analysis samples are positively related to differences in average program-level emphasis on innovation.

Associations with achievement within analysis samples. Commonality analysis for the four sets of predictors in the growth model examined for this report showed that Level of Innovation and Number of Minutes per Day (as a set) uniquely accounted for more than one percent of the variance in CTBS Arithmetic Total Score posttest variance in only three of 21 analysis samples (see Attachment D, Table D-11). In two of these samples (Cohort 0 - SP73 and Cohort 1 - SP73), the regression coefficients were negative for Level of Innovation and positive for Number of Minutes per Day (see Tables 14 and 15). In fact, the majority of the statistically significant regression coefficients for Level of Innovation in all analysis samples, taken as a whole, were negative.

About one-half of the statistically significant regression coefficients for Number of Minutes per Day were positive. Interestingly, positive coefficients tended to be present in Cohorts 0, 1 and 2 and Cohorts 8, 9 and 10. Cohorts 3 through 7 tended to receive negative coefficients. In addition, unlike the reading analysis discussed previously, there was no apparent trend for the coefficients to be positive in the 1971-72 analysis and negative in the 1972-73 analysis.

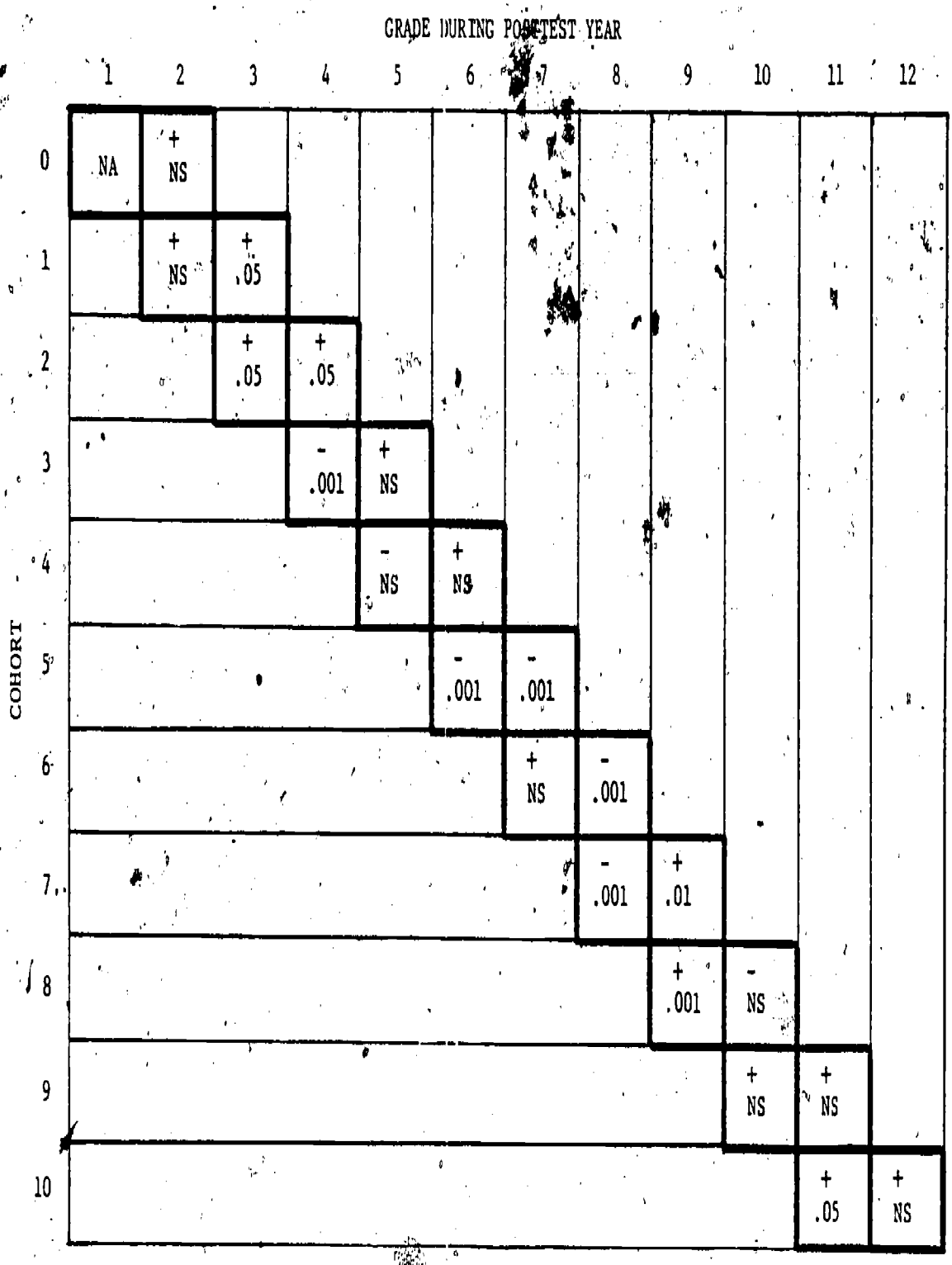
TABLE 14

Sign and Statistical Significance of the Level of Innovation Regression Coefficients - Arithmetic Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR											
	1	2	3	4	5	6	7	8	9	10	11	12
0	NA	- .001										
1		- NS	- .001									
2			- NS	- .05								
3				- .001	+ .05							
4					+ NS	- .01						
5						- .001	- .001					
6							+ .001	- .001				
7								- NS	- .001			
8									- .001	- NS		
9										- NS	+ .05	
10											- .05	+ .05

TABLE 15

Sign and Statistical Significance of the Number of Minutes per Day Regression Coefficients - Arithmetic Total Score Analyses



Although the uniquenesses for Teaching Qualifications were small in all analysis samples (see Attachment D, Table D-12), it was notable that all but one of the statistically significant regression coefficients for this teacher index were positive (Table 16).

Correlations with the residualized posttest score (Tables 17, 18 and 19) also tend to show across samples as follows:

- Level of Innovation was primarily negatively related to the arithmetic posttest residual score.
- Number of Minutes per Day was positively related to the posttest residual score in the young cohorts (Cohorts 0, 1 and 2) and in the oldest cohorts (Cohorts 9 and 10) but negatively associated with the posttest residual in the middle cohorts.
- Teaching Qualifications was primarily positively correlated with the arithmetic posttest residual score.

In review, the regression coefficients and correlations with the posttest residual score both suggest that program-level emphasis on innovation was not highly related to growth in arithmetic achievement and that the predominance of the small associations present were primarily negative. The tendency for Number of Minutes per Day (of typical class time spent on arithmetic/mathematics activities) to be associated negatively with growth in the middle, upper elementary, and junior high school cohorts could not be readily explained.

Comparisons of students who achieved substantially more than expected (for two consecutive school years) with students who achieved less than expected show that, in general, consistent longitudinal overachievement was positively related to Year 1 test level and SES level (see Attachment D, Tables D-14 and D-15). Overachiever/underachiever group differences on these measures, however, were dramatically less than group differences with respect to the Year 2 and Year 3 test scores (see Attachment D, Table D-16). This was a result, of course, of the procedure used to identify these particular students. Findings tended to parallel those just summarized for the regression analyses and residual correlations (see Attachment

TABLE 16

Sign and Statistical Significance of the Teaching Qualifications Regression Coefficients - Arithmetic Total Score Analyses

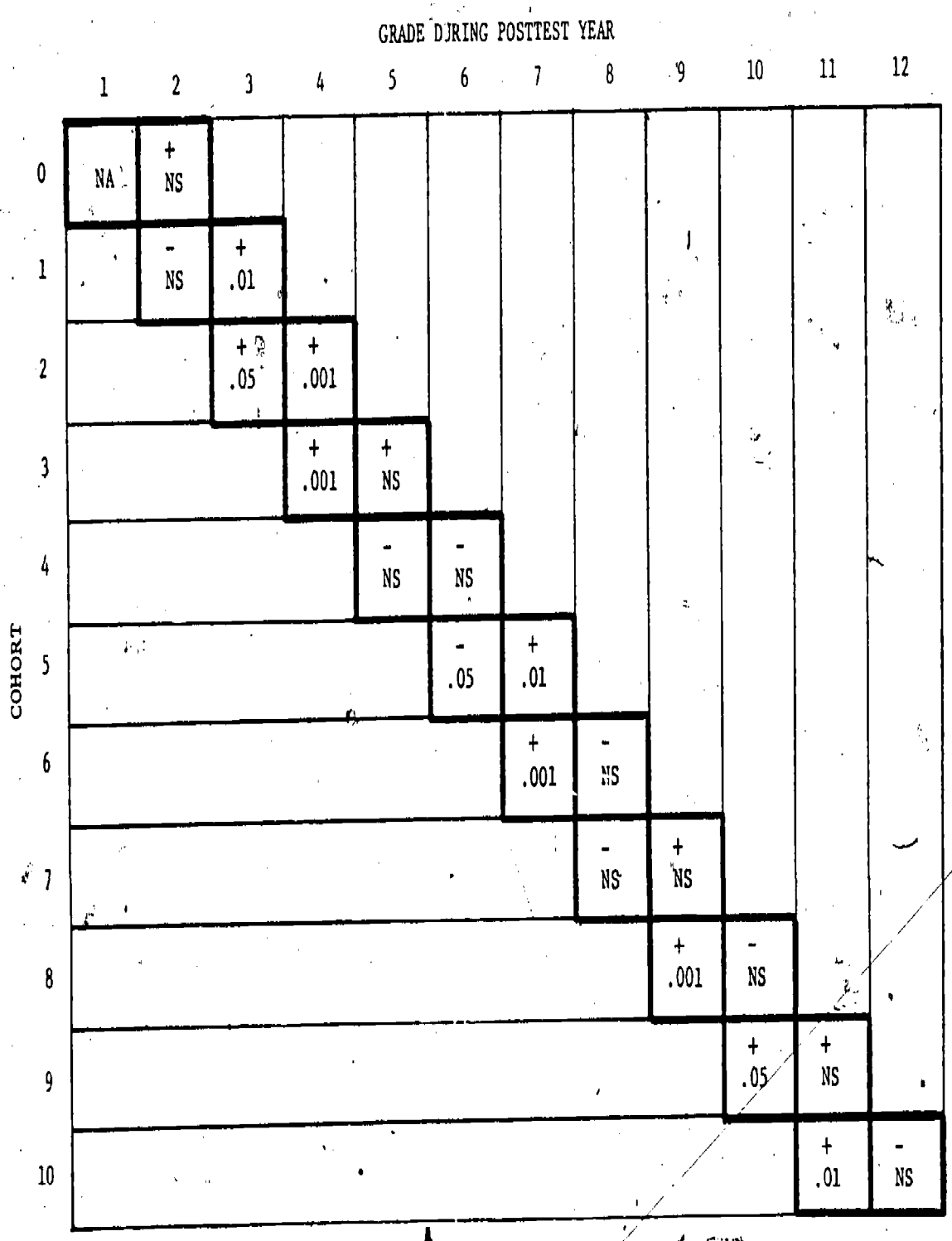


TABLE 17

Correlations Between a Residualized Posttest Score and Level of Innovation - Arithmetic Total Score Analyses

COHORT	GRADE DJRING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	-.210										
1		-.061	-.275									
2			-.065	-.094								
3				-.070	.051							
4					.016	-.060						
5						-.114	-.092					
6							.129	-.126				
7								-.071	-.031			
8									-.046	-.036		
9										.035	.174	
10											-.052	.238

TABLE 18

Correlations Between a Residualized Posttest Score and Number of Minutes per Day - Arithmetic Total Score Analyses

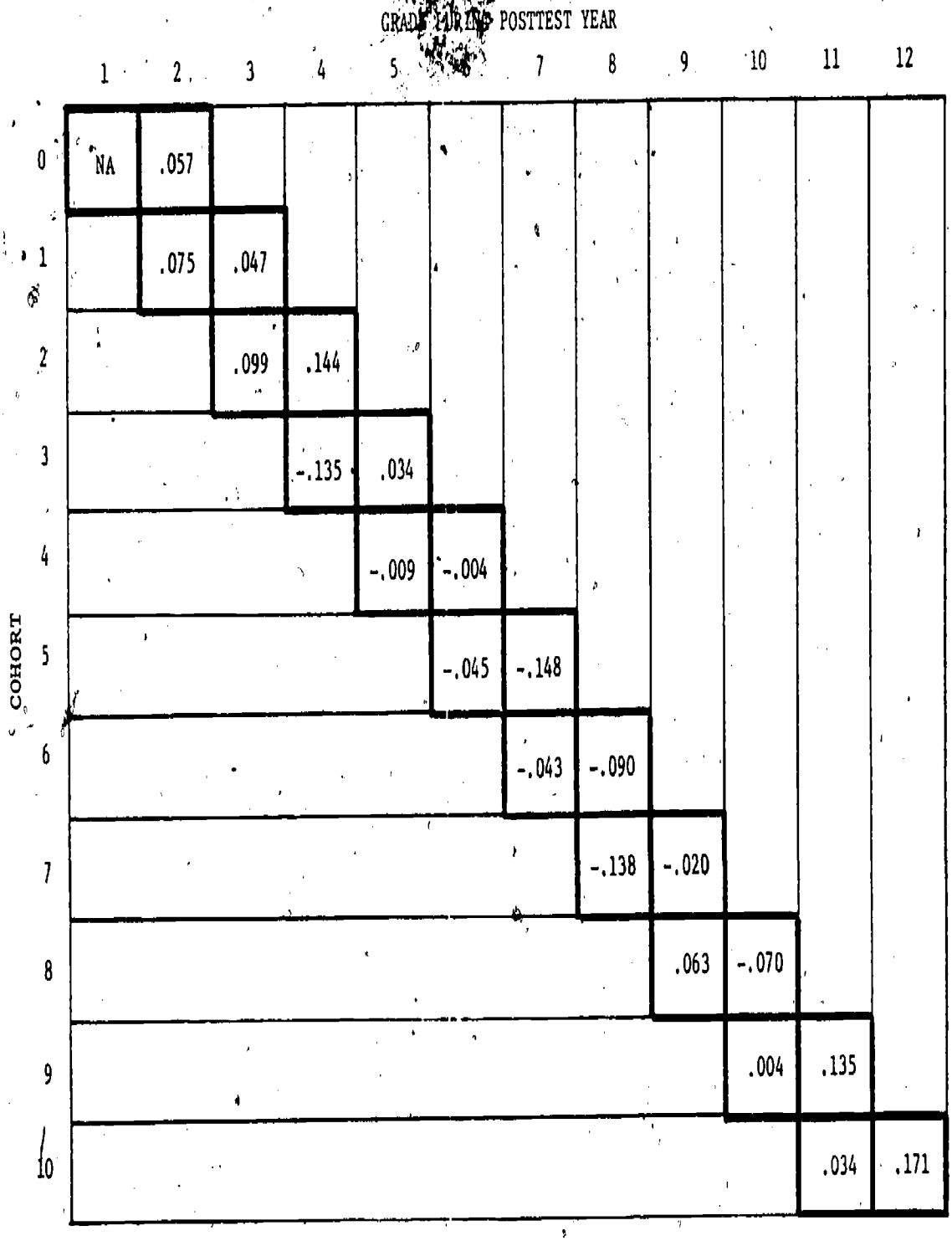
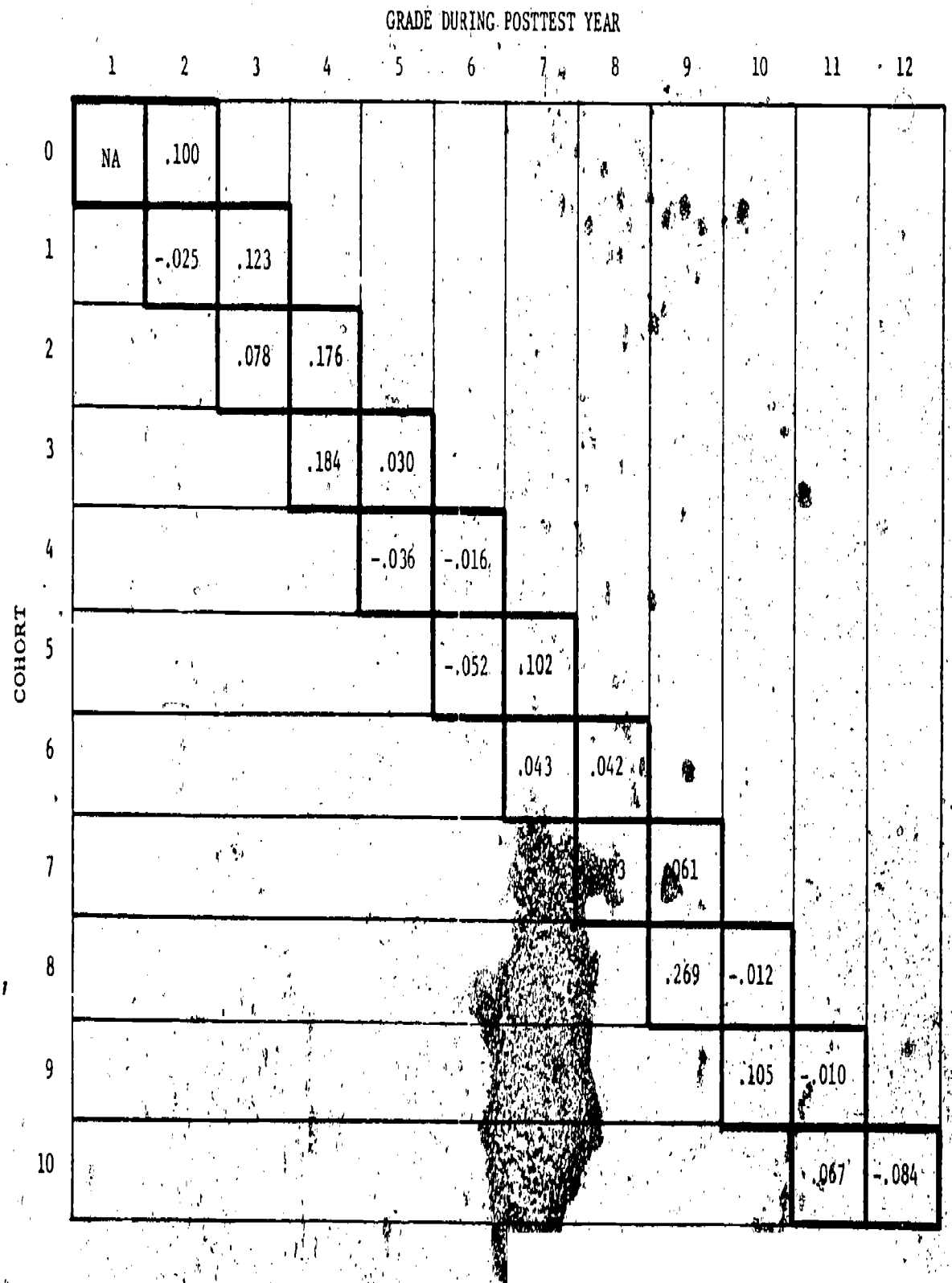


TABLE 19

Correlations Between a Residualized Posttest Score and Teaching Qualifications - Arithmetic Total Score Analyses



D, Tables D-17, D-18 and D-19. Relative to the underachievers, over-achievers tended, on the average,

- to be members of programs with less emphasis on innovation during the 1972-73 school year in Cohorts 1 through 8
- to be members of programs with greater emphasis on innovation in Cohorts 9 and 10
- to be exposed to less time on arithmetic class activities in Cohorts 3 through 8 but more time on arithmetic in Cohorts 1 and 2 and Cohorts 9 and 10
- to be taught by arithmetic/mathematics teachers who obtained higher scores on Project LONGSTEP's measure of teacher experience.

The within-analysis-sample findings for arithmetic achievement, then, show that

- Level of Innovation, in general, was not highly associated with arithmetic achievement and that there was a somewhat consistent tendency for slightly greater arithmetic achievement to be associated with less program-level emphasis on innovation
- Number of Minutes per Day was positively related to arithmetic achievement but only during early elementary school and late high school
- growth in arithmetic achievement was positively associated with Teaching Qualifications.

Overview of Trends for Reading and Arithmetic Achievement

The purpose of the analyses conducted for this supplement to Volume I of the Project LONGSTEP final report was (1) to ascertain if the trends observed and discussed in Volume I with respect to Cohorts 1, 4 and 6 were representative of the trends shown by all analysis samples and (2) to compare results and determine if other meaningful trends across cohorts were present. The overall findings reported here have shown the following:

- The mean reading and arithmetic posttest scores for Project LONGSTEP's sample of fairly innovative schools were not conspicuously farther from national norms than their average pretest scores were from their norms.
- Variations among analysis samples with respect to average reading and arithmetic achievement gains did not tend to be associated in any highly consistent manner with concomitant differences in the mean Level of Innovation, Number of Minutes per Day, and Teaching Qualifications of the samples.
- Variation in Level of Innovation was not highly associated with reading or arithmetic achievement within Project LONGSTEP's analysis samples.
- Variation in Level of Innovation was not consistently or positively related to reading achievement within analysis samples.
- Variation in Level of Innovation appeared to be negatively associated with arithmetic achievement.
- Teaching Qualifications was not highly or consistently related to reading achievement but was positively associated with small gains in arithmetic achievement.

Additional findings included (1) a trend for reading and arithmetic achievement to decrease between the sixth and seventh grades, (2) a trend for reading and arithmetic gains to be larger in the earlier grades, (3) a trend for the average gains in arithmetic shown by all cohorts except those in senior high to be larger than mean gains in reading (relative to their respective standard deviations), (4) a trend for the elementary grades (1-6) to be exposed to notably more instruction per day than junior high and high school students with respect to language arts, and (5) notable mean achievement gains in reading and arithmetic for students who were third-graders during the 1972-73 school year.

In summary, the primary research hypothesis, that substantial gains in cognitive achievement are positively associated with innovative emphasis, has not been supported in any general way by the analysis of Project LONGSTEP's data. These results, based on a global analysis of trends across reading and arithmetic analysis samples and cohorts/grades, tend to confirm the findings reported in Volume I for Cohorts 1, 4 and 6.

REFERENCES

Coles, G. J., Chalupsky, A. B., Everett, B. E., Shaycoft, M. F., Rodabaugh, B. J., & Danoff, M. N. Impact of educational innovation on student performance: Project methods and findings for three cohorts. Project LONGSTEP final report: Volume I. Palo Alto, California: American Institutes for Research, 1976.

U. S. Department of Health, Education, and Welfare, Office of Education. A procedural guide for validating achievement gains in educational projects. Washington, D. C.: U. S. Government Printing Office, 1976.

ATTACHMENTS

ATTACHMENT A

Location of the Complete Statistical Tables Summarized in Volume I Supplement

Analytic methods implemented specifically for the Volume I Supplement involved the data for Cohorts 0, 2, 3, 5, 7, 8, 9, and 10 for two cognitive dependent variables, the CTBS Reading Total and Arithmetic Total Scores. The focus of this report, however, was not on detailed comparisons of the results shown by individual cohorts but on the overall conclusions based on findings from all cohorts (including Cohorts 1, 4, and 6 discussed in Volume I). Thus, the cohort-by-grade tables incorporated in the Volume I Supplement summarize results obtained either in these analyses or in those conducted for Volume I itself. Table A-1 shows the location of the statistical analysis tables for each cohort. The content of the Volume I Supplement Appendices is shown in Table A-2

Table A-1

Volume Location* of All Statistical Analysis Tables

Cohort	Volume I or Volume I Appendices	Volume I Supplement Appendices
0		X
1	X	
2		X
3		X
4	X	
5		X
6	X	
7		X
8		X
9		X
10		X

*Volume I, the Volume I Appendices, and the Volume I Supplement Appendices are each separately bound documents and will be deposited in ERIC at the conclusion of the project.

TABLE A-2

Contents of the Volume I
Supplement Appendices

Appendix	Content	
A	All Reading Analysis Tables for Cohorts 0, 2, 3, 5, 7, 8, 9, and 10	
	Tables	
	A-1 to A-15	Means and Standard Deviations
	B-1 to B-15	Intercorrelations
	C-1 to C-15	Commonality Analysis Tables
	D-1 to D-2	Regression Analysis Results
B	All Arithmetic Analysis Tables for Cohorts 0, 2, 3, 5, 7, 8, 9, and 10	
	Tables	
	A-1 to A-15	Means and Standard Deviations
	B-1 to B-15	Intercorrelations
	C-1 to C-15	Commonality Analysis Tables
	D-1 to D-2	Regression Analysis Results
C	Project LONGSTEP Profile Variables and Variable Abbreviation Summary	

ATTACHMENT B

Supporting Tables* - Reading Total Score Analyses

*The first box in each row contains results for the Spring 1972 posttest analysis; the second box contains the Spring 1973 posttest analysis results. In boxes containing two values, the first statistic described in the table title is the upper value and the second statistic described is the lower value. Cells along the diagonal for which data were "not available" or "not applicable" contain an "NA." The abbreviation "NS" has been used for "non-significant."

TABLE B-1

Pretest National Norms (50th Percentile) - Reading Total Score Analyses

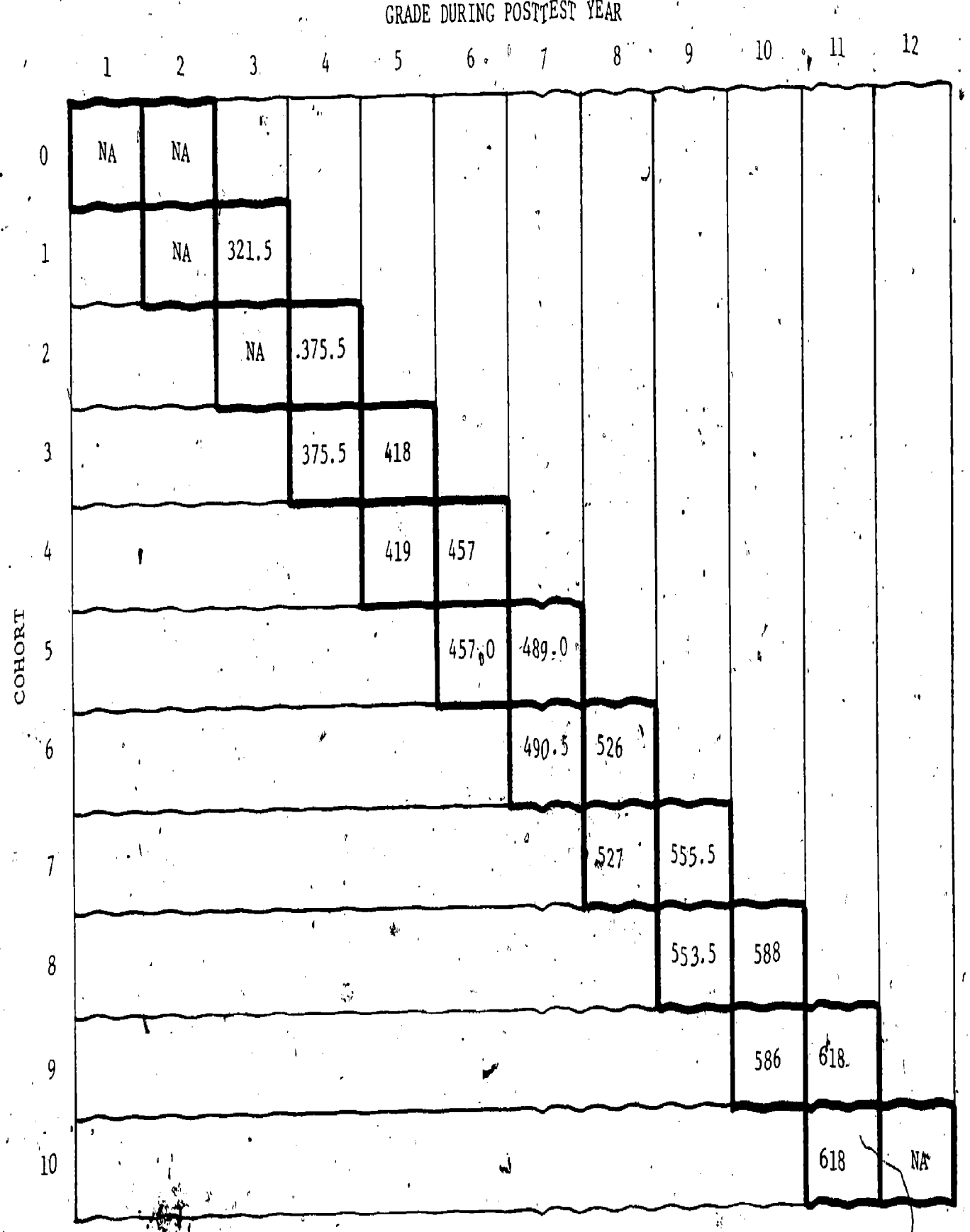


TABLE B-2

Posttest National Norms (50th Percentile) - Reading Total Score Analyses

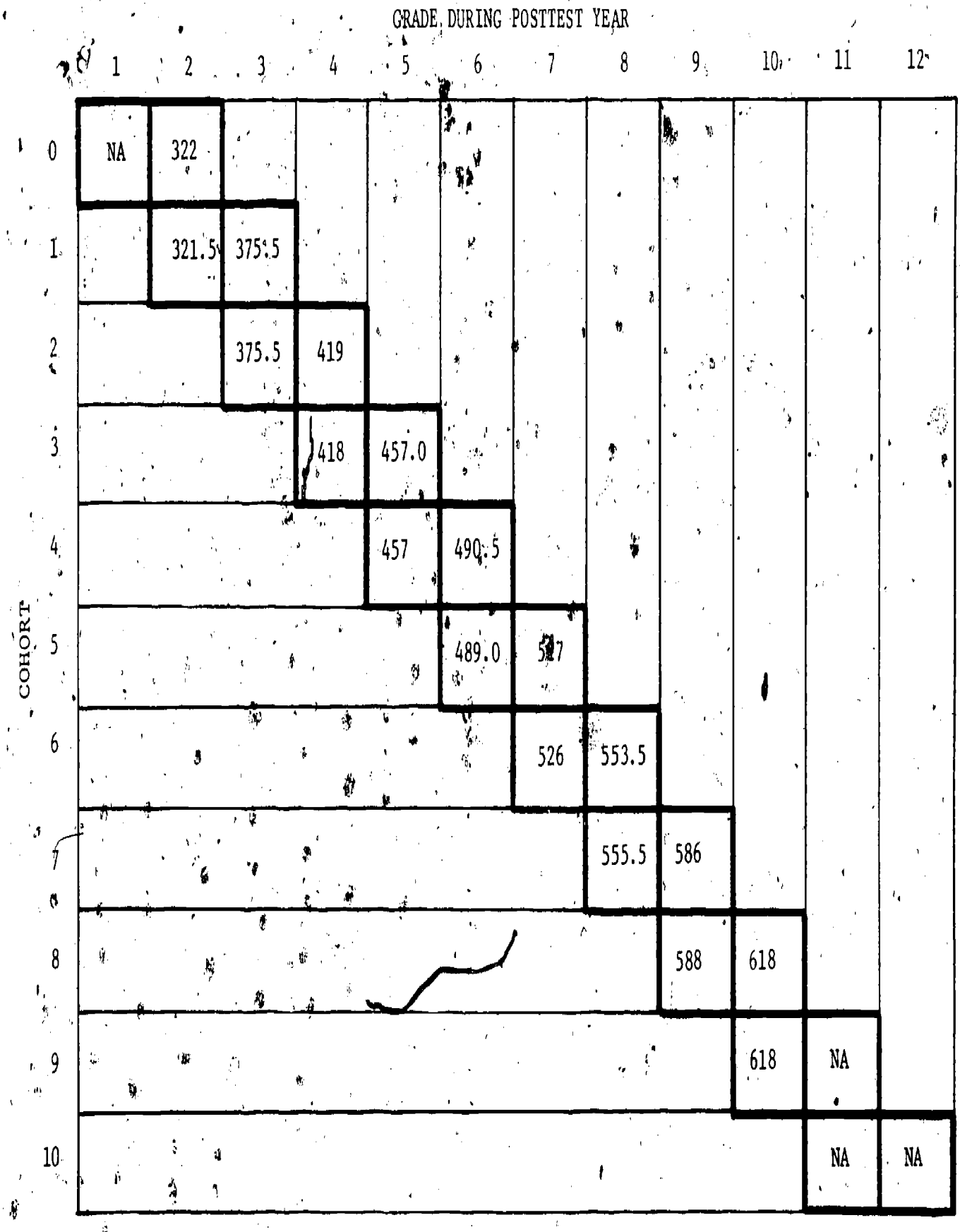


TABLE B-3

Number of Students by Analysis Sample - Reading Total Score Analyses

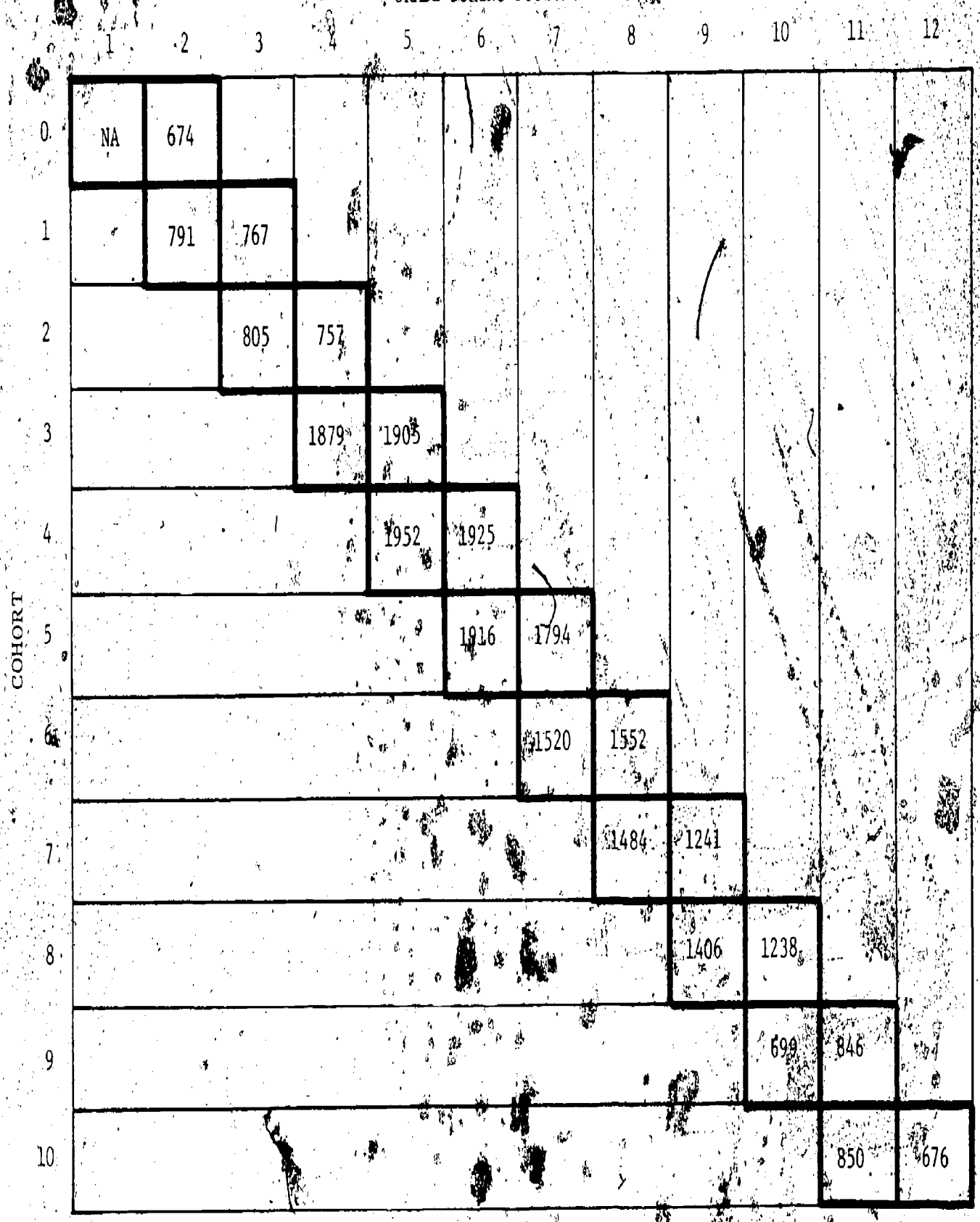


TABLE B-4

Pretest Means and Standard Deviations - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	92.44* 11.00										
1		93.85* 10.56	338.46 63.87									
2			105.28* 11.36	386.11 63.49								
3				390.04 63.72	435.07 60.56							
4					437.91 64.25	475.73 67.30						
5						473.82 71.00	512.31 65.08					
6							513.30 72.78	545.45 72.79				
7								538.00 72.43	580.83 70.68			
8									579.25 73.14	616.88 75.81		
9										619.84 83.04	652.38 72.64	
10											668.06 82.23	688.12 76.46

*Based on the CTMM.

TABLE B-5

Socioeconomic Status (SES) Means and Standard Deviations - Reading Total Score Analyses

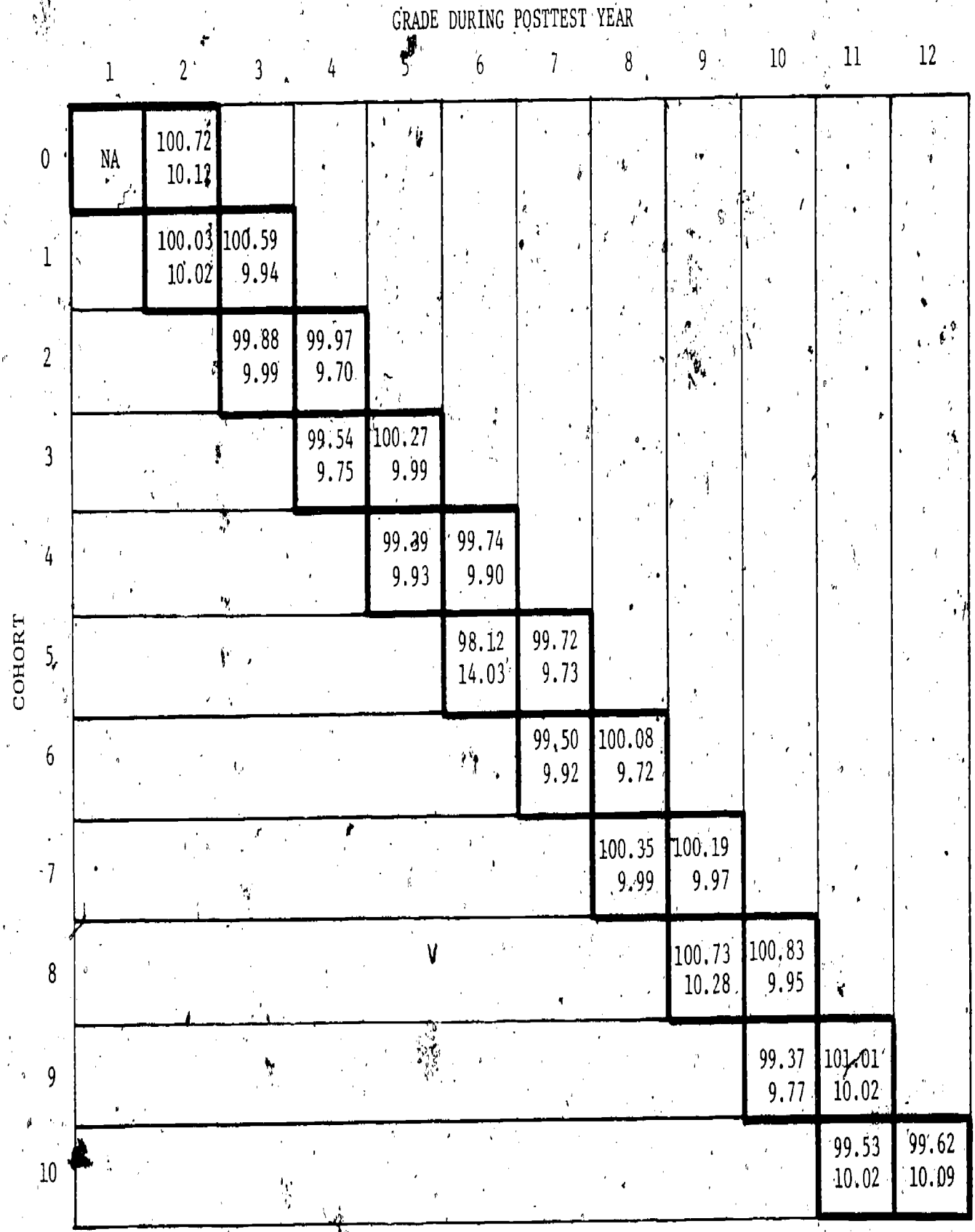


TABLE B-6

Level of Innovation Means and Standard Deviations - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	21.25 3.07										
1		20.12 4.33	21.03 3.26									
2			20.12 4.06	20.86 3.19								
3				18.50 4.23	19.51 3.32							
4					18.77 4.43	19.95 3.55						
5						19.37 3.91	18.35 4.49					
6							19.24 4.74	19.25 4.46				
7								19.65 4.33	17.49 3.61			
8									18.18 3.69	18.01 3.61		
9										20.15 4.62	17.74 3.13	
10											17.73 3.54	18.12 3.39

TABLE B-7

Teaching Qualifications Means and Standard Deviations - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	97.87 7.50										
1		97.61 7.84	100.90 6.35									
2			97.90 9.78	100.53 7.19								
3				101.18 7.65	100.67 8.05							
4					102.47 7.88	99.86 7.46						
5						100.12 8.60	98.84 7.23					
6							95.26 8.24	99.06 6.53				
7								97.47 9.17	101.63 8.27			
8									100.56 8.47	101.01 7.70		
9										99.44 7.20	100.72 8.80	
10											104.74 8.13	103.24 9.39

TABLE B-8

School Membership by Grade and Cohort[1]

School	Cohort 0	1	2	3	4	5	6	7	8	9	10	Site
1				4,5	5,6	6,-						
2				4,5	5,6	6,-						A
3						-,7						
4				4,5	5,6	6,-						
5						-,7						B
6				4,5	5,6	6,-						
7				4,5	5,6	6,-						
8				4,5	5,6	6,-						
9				4,5	5,6	6,-						
10				4,5	5,6	6,-						C
11				4,5	5,6	6,-						
12				4,5	5,6	6,-						
13						-,7						
14						-,7	7,8	8,9	9,-			
15									-,10	10,11	11,12	
16	1,2	2,3	3,4	4,5	5,-							
17	1,2	2,3	3,4	4,5	5,-							
18					-,6	6,7	7,8	8,-				D
19								-,9	9,10	10,11	11,12	
20				4,5	5,6	6,-						
21						-,7						E
22				4,5	5,6	6,-						
23				4,5	5,6	6,-						
24				4,5	5,6	6,-						F
25						-,7	7,8	8,-				
26								-,9	9,10			
27				4,-[2]								
28				4,5	5,-							
29				4,-[2]								
30				4,-[2]								G
31					-,6	6,7	7,8	8,-				
32								-,9	9,10			

(continued)

TABLE B-8 (continued)

School	Cohort 0	1	2	3	4	5	6	7	8	9	10	Site
33				4,5	5,-	6,-						
34	1,2	2,3	3,4	4,5	5,6	6,-						
35						-,7						H
36					-,6	-,7						
37				4,5	5,6	6,-						
38				4,5	5,6	6,-						
39						-,7	7,8	8,9	9,-			I
40									-,10	10,11	11,12	
41	1,2	2,3	3,4	4,5	5,6	6,-						
42	1,2	2,3	3,4	4,5	5,6	6,-						
43	1,2	2,3	3,4	4,5	5,6	6,-						
44	1,2	2,3	3,4	4,5	5,6	6,-						J
45						-,7						
46						-,7						
47	1,2	2,3	3,4	4,5	5,6	6,-						
48	1,2	2,3	3,4	4,5	5,6	6,-						
49						-,7	7,8	8,9	9,-			K
50						-,7						
51									-,10			
52	1,2	2,3	3,4	4,5	5,6	6,-						
53	1,2	2,3	3,4	4,5	5,6	6,-						
54	1,2	2,3	3,4	4,5	5,6	6,-						L
55	1,2	2,3	3,4	4,5	5,6	6,-						
56						-,7	7,8	8,-				
57								-,9				
58				4,5	5,6	6,-						
59				4,5	5,6	6,7						
60				4,5	5,6	6,7						M
61				4,5	5,6	6,-						
62						-,7						

[1]Table entries in each column are the grades for the cohort in 1971-72 (first number) and 1972-73 (second number).

[2]School closed and students transferred.

TABLE B-9

Uniquenesses and Multiple Correlations (Squared) for Pretest - Reading Total Score Analyses

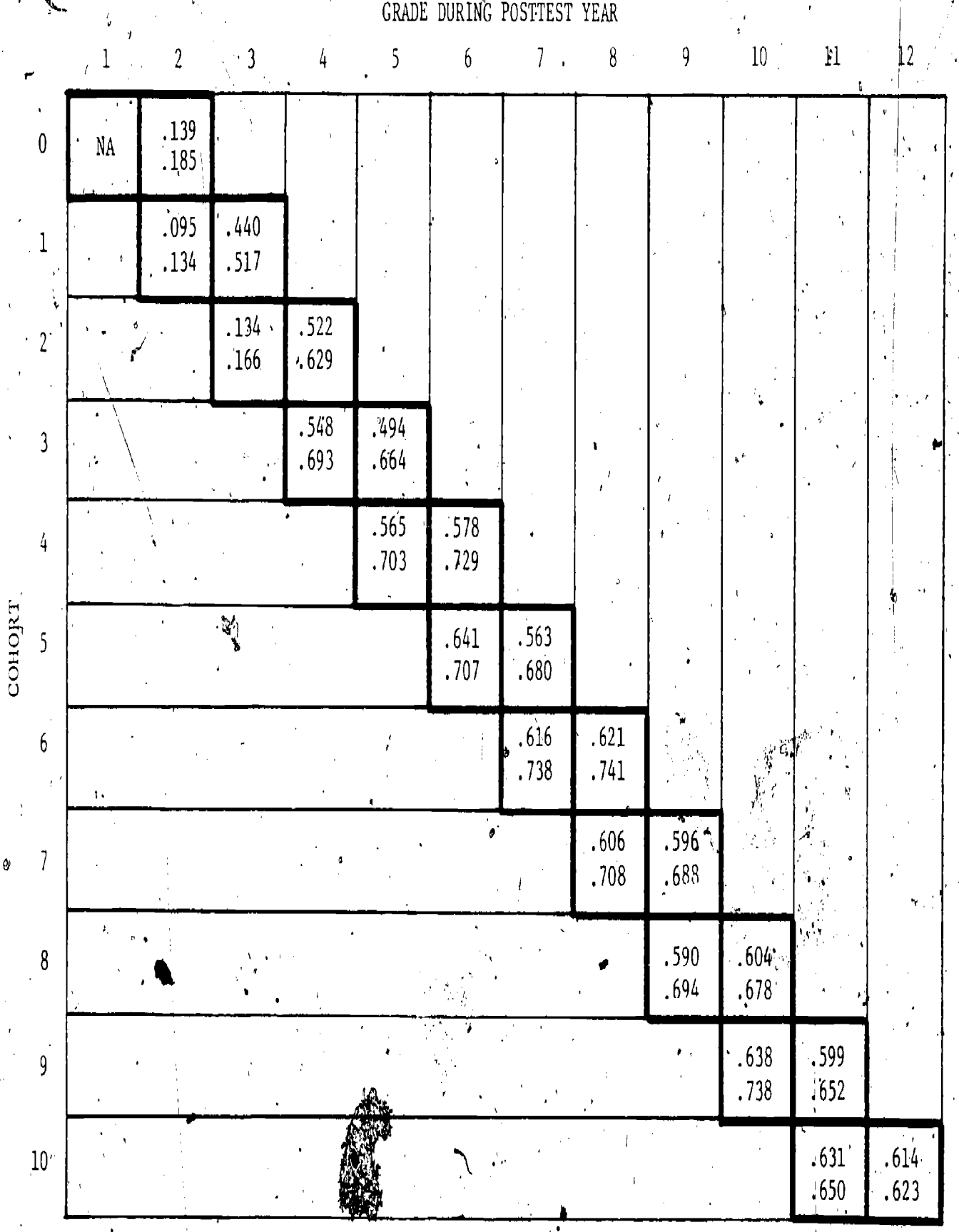


TABLE B-10

Uniquenesses and Multiple Correlations (Squared) for SES - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	.021 .072										
1		.114 .164	.001 .075									
2			.073 .112	.008 .097								
3				.005 .149	.013 .182							
4					.008 .138	.006 .152						
5						.001 .062	.004 .116					
6							.005 .114	.001 .118				
7								.004 .102	.001 .090			
8									.001 .104	.002 .085		
9										.002 .105	.002 .045	
10											.001 .017	.000 .004

TABLE B-11

Uniquenesses and Multiple Correlations (Squared)
for Level of Innovation and Number of Minutes per Day - Reading Total Score Analyses

GRADE DURING POSTTEST YEAR

COHORT	1	2	3	4	5	6	7	8	9	10	11	12
0	NA.	.013 .009										
1		.004 .008	.088 .047									
2			.005 .002	.005 .008								
3				.004 .012	.002 .009							
4					.002 .002	.001 .016						
5						.000 .004	.000 .021					
6							.003 .045	.007 .045				
7								.002 .028	.010 .016			
8									.003 .006	.021 .008		
9										.008 .024	.016 .001	
10											.001 .003	.013 .018

TABLE B-12

Uniquenesses and Multiple Correlations (Squared)
for Teaching Qualifications - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	.001 .000										
1		.002 .000	.006 .040									
2			.012 .008	.001 .000								
3				.001 .000	.001 .001							
4					.000 .003	.000 .000						
5						.000 .000	.002 .012					
6							.000 .025	.000 .017				
7								.000 .016	.001 .001			
8									.002 .017	.000 .000		
9										.000 .000	.000 .017	
10											.000 .006	.003 .029

TABLE B-13

Numbers of Consistently Overachieving and Underachieving Students Identified - Reading Total Score Analyses

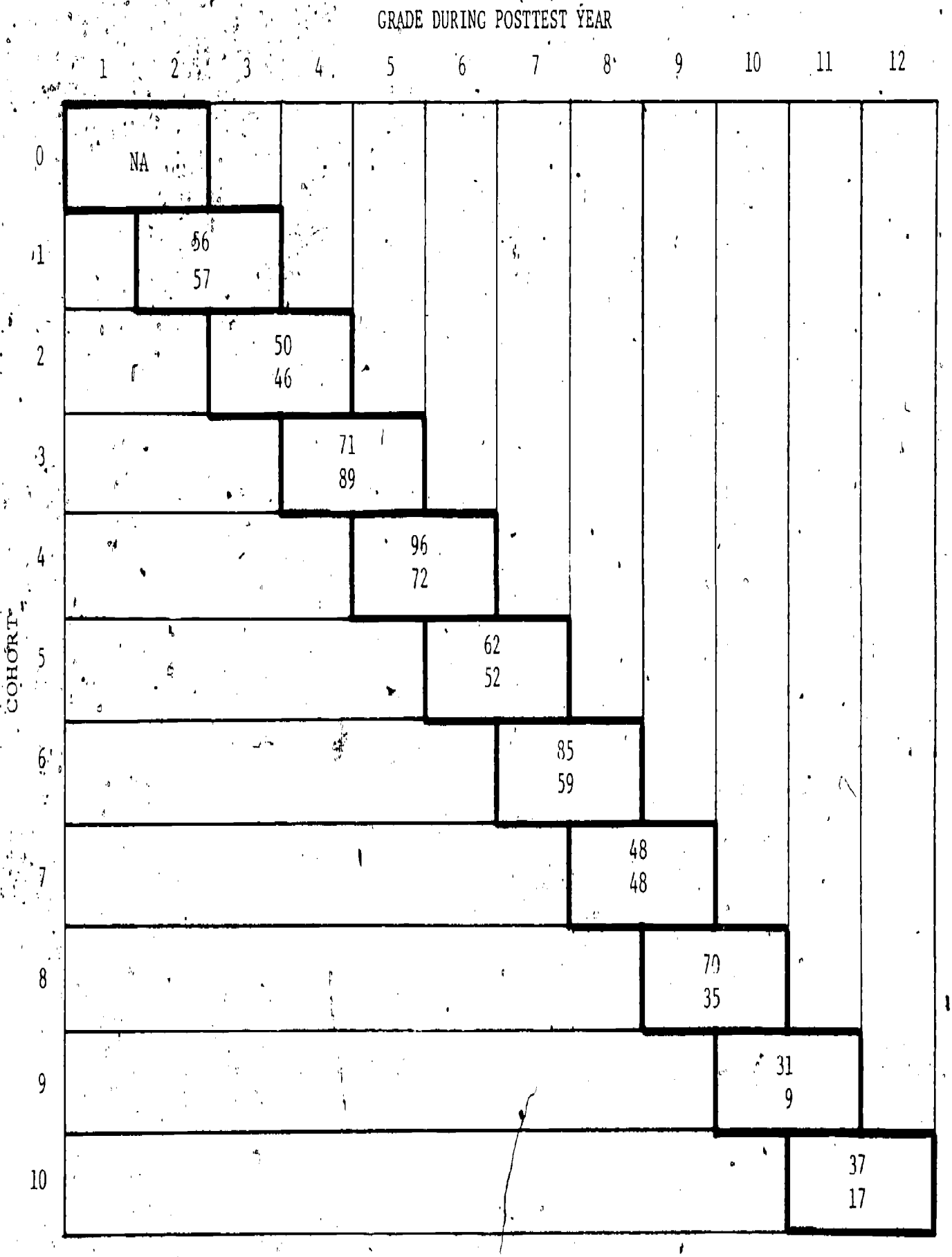


TABLE B-14

Point-Biserial Correlations Between Student Outlier Group Membership and Pretest - Reading Total Score Analyses

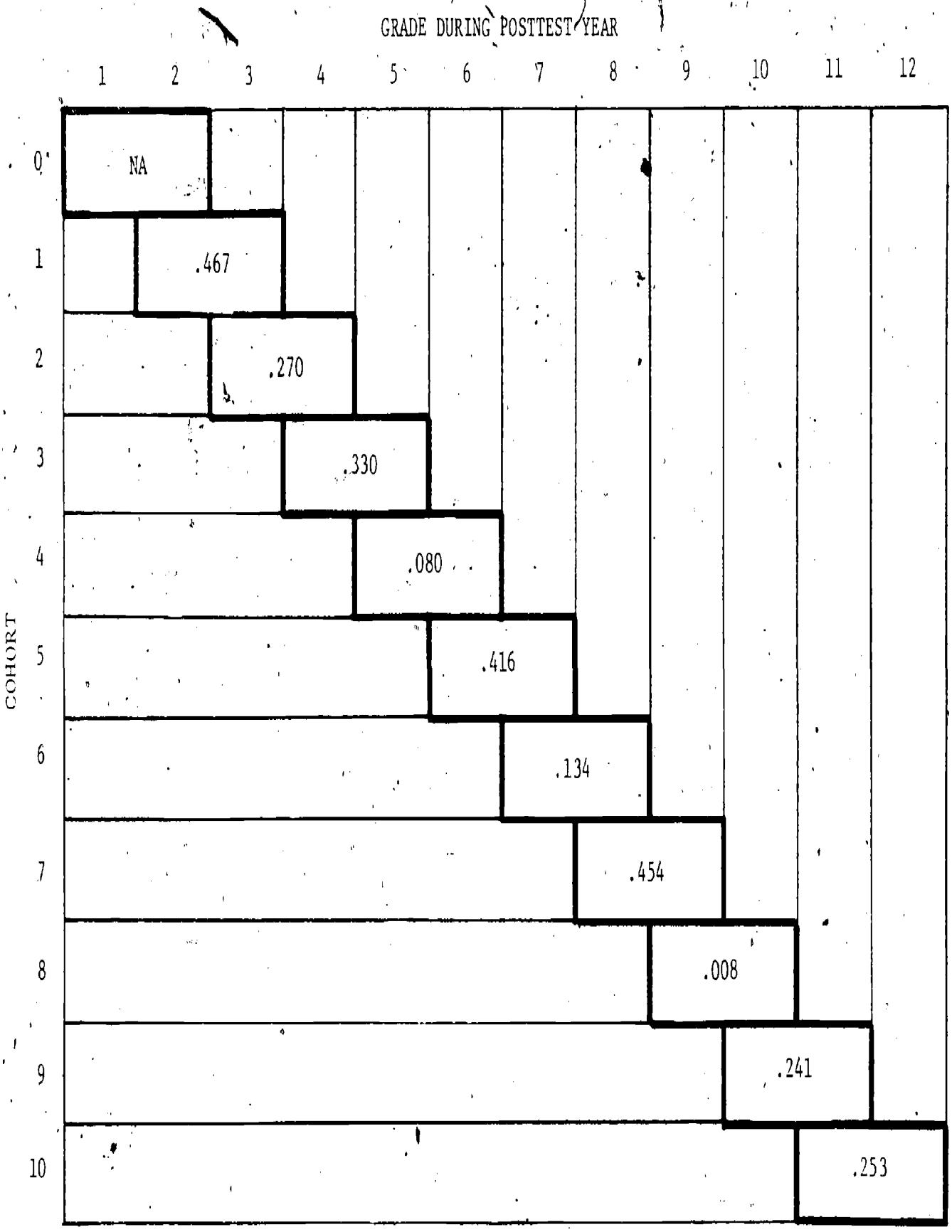


TABLE B-15

Point-Biserial Correlations Between Student Outlier Group Membership and SES - Reading Total Score Analyses

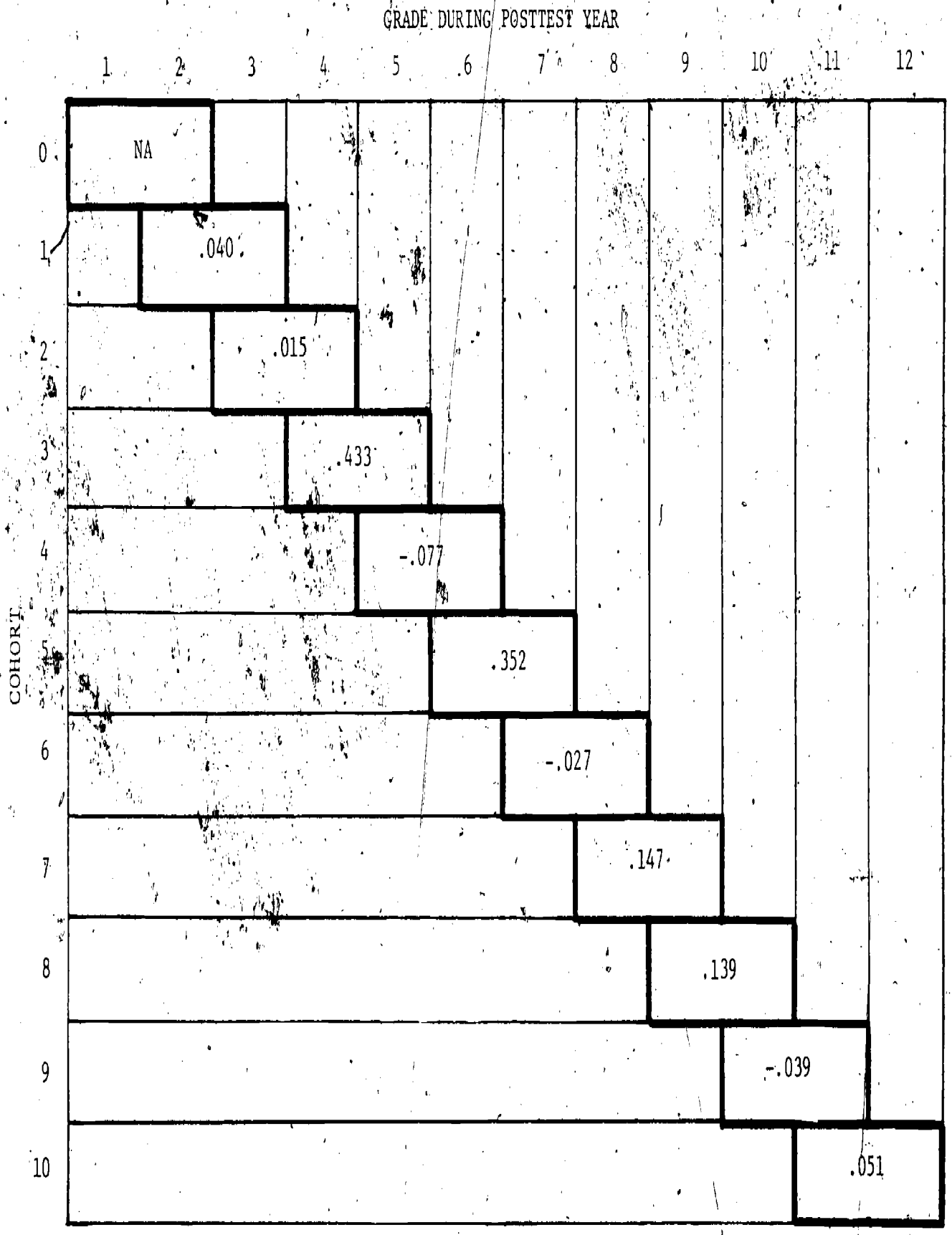


TABLE B-16

Point-Biserial Correlations Between Student Outlier Group Membership and Year 2 and Year 3 Test Score - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	NA										
1		.894	.943									
2			.895	.942								
3				.704	.854							
4					.621	.813						
5						.728	.861					
6							.591	.794				
7								.778	.891			
8									.512	.761		
9											833	
10											.700	.857

TABLE B-17

Point-Biserial Correlations Between Student Outlier Group Membership and Level of Innovation - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	NA										
1		-.170	-.438									
2			-.132	-.041								
3				.017	-.129							
4					-.161	-.041						
5						-.101	-.116					
6							-.047	-.102				
7								.144	-.068			
8									-.145	.249		
9										.341	.208	
10											.464	.465

TABLE B-18

Point-Biserial Correlations Between Student Outlier Group Membership and Number of Minutes per Day - Reading Total Score Analyses

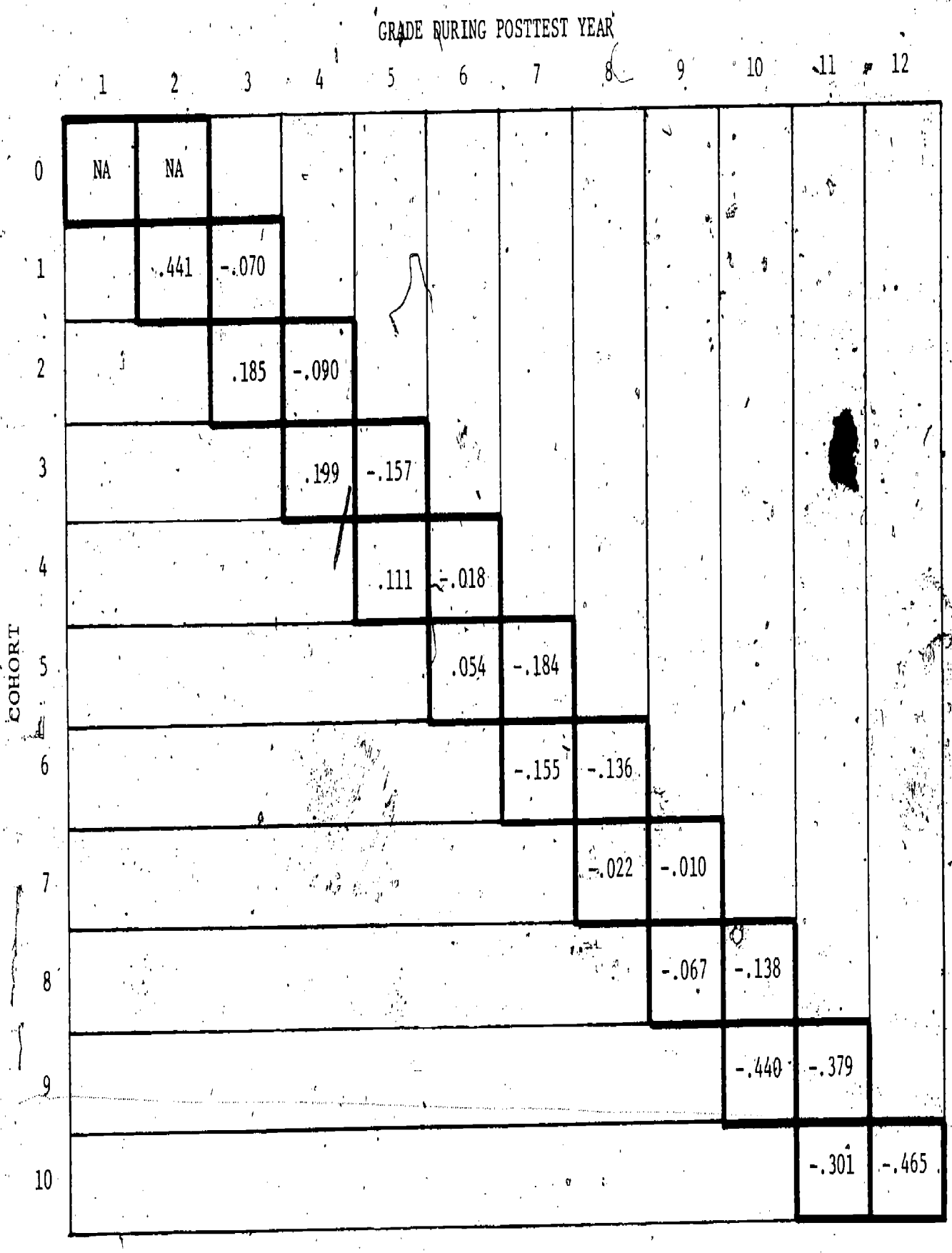


TABLE B-19

Point-Biserial Correlations Between Student Outlier Group Membership and Teaching Qualifications - Reading Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	NA										
1		.041	.333									
2			.081	.069								
3				-.087	.041							
4					.161	-.015						
5						.050	-.155					
6							-.077	.129				
7								.061	-.061			
8									-.034	-.043		
9										-.155	.051	
10											.202	.468

ATTACHMENT C

Notes on Correlations with
Residualized Posttest Scores[1]

A tendency was noted across samples for the correlation of Number of Minutes per Day during Year 2 to be positively correlated with the Year 2 reading posttest residual score and for Number of Minutes per Day in Year 3 to be negatively correlated with the Year 3 reading posttest residual. Project staff considered that this seemingly consistent trend in the reading analyses might have been an artifact completely due to use of the Year 2 test as both a posttest (in the Year 2 posttest analyses) and as a pretest (in the Year 3 posttest analyses). This attachment notes a set of circumstances that could lead to this trend but does not conclude that the shift in sign is, by necessity, a statistical artifact.

For the sake of simplicity, let us restrict our example to a posttest score residualized on the basis of a single predictor, pretest. In this case, the Year 2 posttest residual can be shown to be equal to

$$Y'_2 = Y_2 - a_1 - b_1 Y_1 \qquad \text{(C-1)}$$

where,

Y'_2 = Year 2 test score residual,

Y_2 = Year 2 test score,

a_1 = raw score intercept for the regression of Y_2 on Y_1,

b_1 = raw score regression coefficient for the regression of Y_2 on Y_1, and

Y_1 = Year 1 test score.

The corresponding residual for the Year 3 posttest test score is,

$$Y'_3 = Y_3 - a_2 - b_2 Y_2 \qquad \text{(C-2)}$$

---

[1] Appreciation is expressed to Dr. Robert Linn who provided the general formulas upon which this attachment is based.

where,

Y'_3 = Year 3 test score residual,

Y_3 = Year 3 test score,

a_2 = raw score intercept for the regression of Y_3 on Y_2,

b_2 = raw score regression coefficient for the regression of Y_3 on Y_2, and

Y_2 = Year 2 test score.

If we now assume that students are exposed to the same Number of Minutes per Day (NMIN) during both Year 2 and Year 3 (an assumption definitely not present in the LONGSTEP data), the covariance of the Year 2 residual, Y'_2, with NMIN will be equal to

$$\text{Cov}\ (Y_2, \text{NMIN}) - b_1 \cdot \text{Cov}\ (Y_1, \text{NMIN}). \qquad \text{(C-3)}$$

Similarly, the corresponding covariance of NMIN with the Year 3 residual, Y'_3, will equal

$$\text{Cov}\ (Y_3, \text{NMIN}) - b_2 \cdot \text{Cov}\ (Y_2, \text{NMIN}). \qquad \text{(C-4)}$$

Now, given that Cov (Y_2, NMIN) is the largest of the three covariance terms and that b_2 is near 1.0, the covariance of NMIN with the Year 2 residual (C-3) will be positive and the covariance with the Year 3 residual (C-4) will be negative. Conversely, if Cov (Y_2, NMIN) is the smallest of the three covariances and b_1 is near 1.0, then the covariance of NMIN with the Year 2 residual (C-3) will be negative and the covariance with the Year 3 residual (C-4) will be positive. Since there is no particular statistical reason for NMIN to be more highly related to the Year 2 or Year 3 test score and since the Number of Minutes per Day index used in these analyses were gathered during Years 2 and 3 separately and did vary between years, the shift in signs observed in the Project LONGSTEP residual correlations is believed to be merely a methodological artifact that is independent the processes being studied.

ATTACHMENT D

Supporting Tables* - Arithmetic Total Score Analyses

*The first box in each row contains results for the Spring 1972 posttest analysis; the second box contains the Spring 1973 posttest analysis results. In boxes containing two values, the first statistic described in the table title is the upper value and the second statistic described is the lower value. Cells along the diagonal for which data were "not available" or "not applicable" contain an "NA." The abbreviation "NS" has been used for "non-significant."

TABLE D-1

Pretest National Norms (50th Percentile) - Arithmetic Total Score Analyses

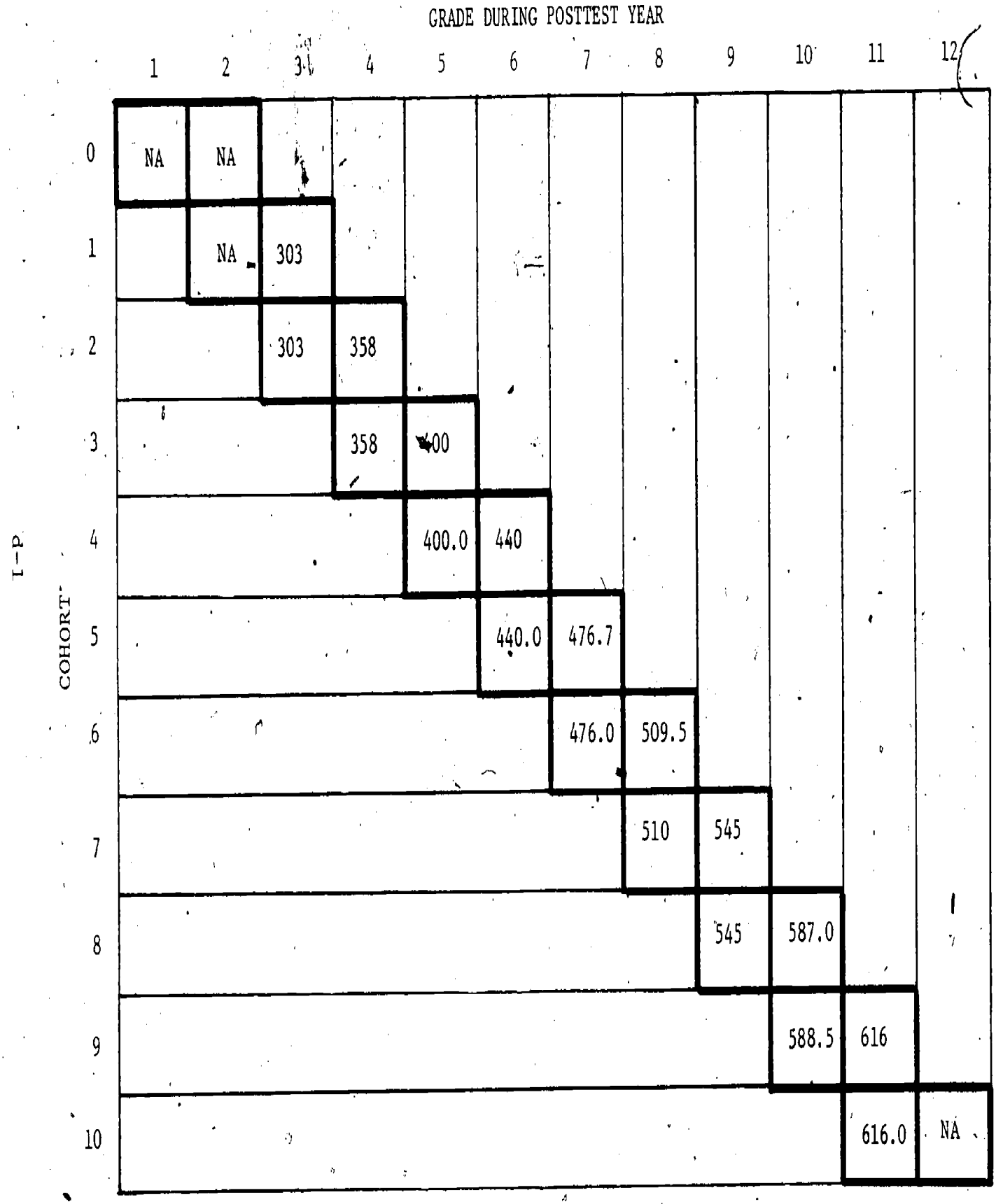


TABLE D-2

Posttest National Norms (50th Percentile) - Arithmetic Total Score Analyses

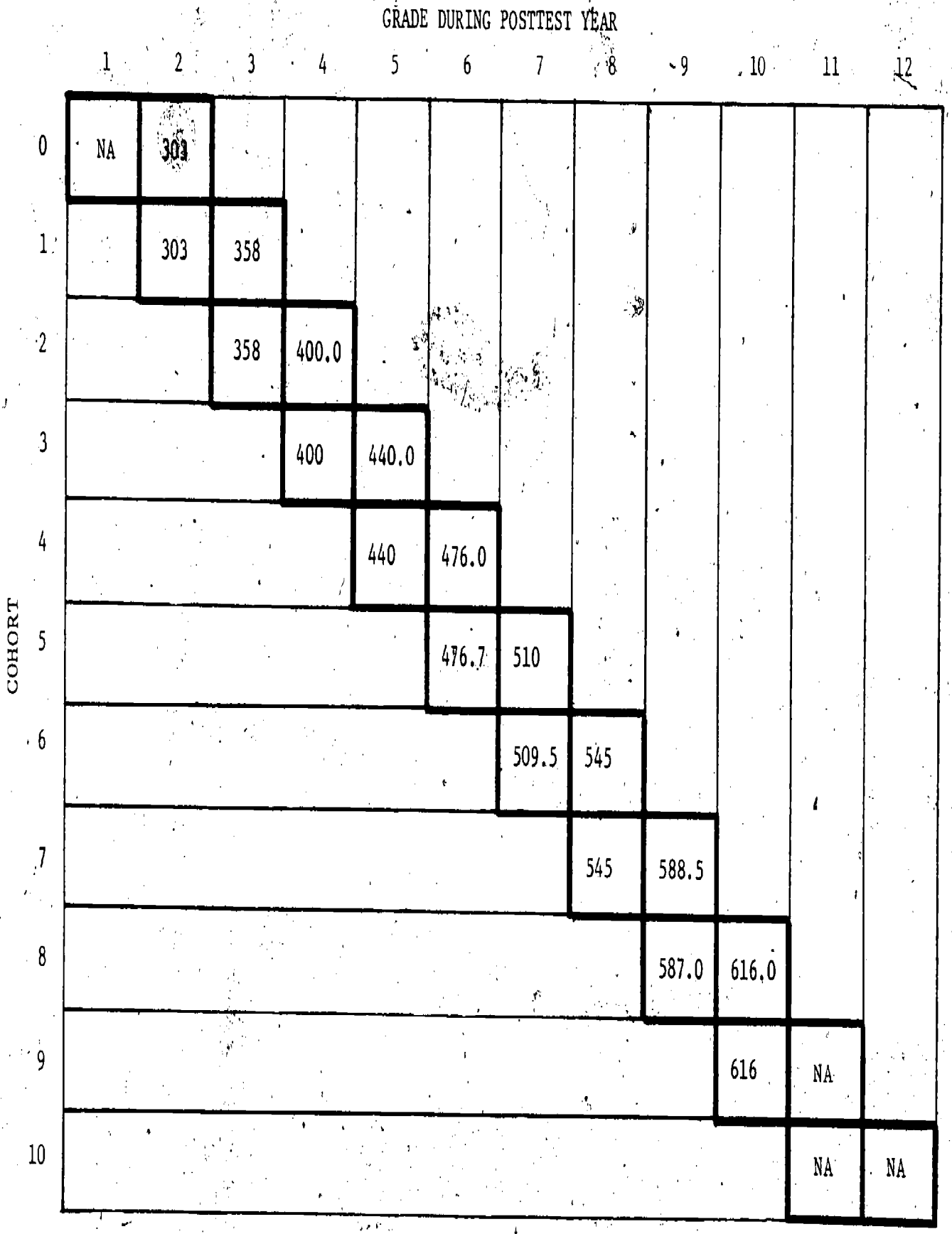


TABLE D-3

Number of Students by Analysis Sample - Arithmetic Total Score Analyses

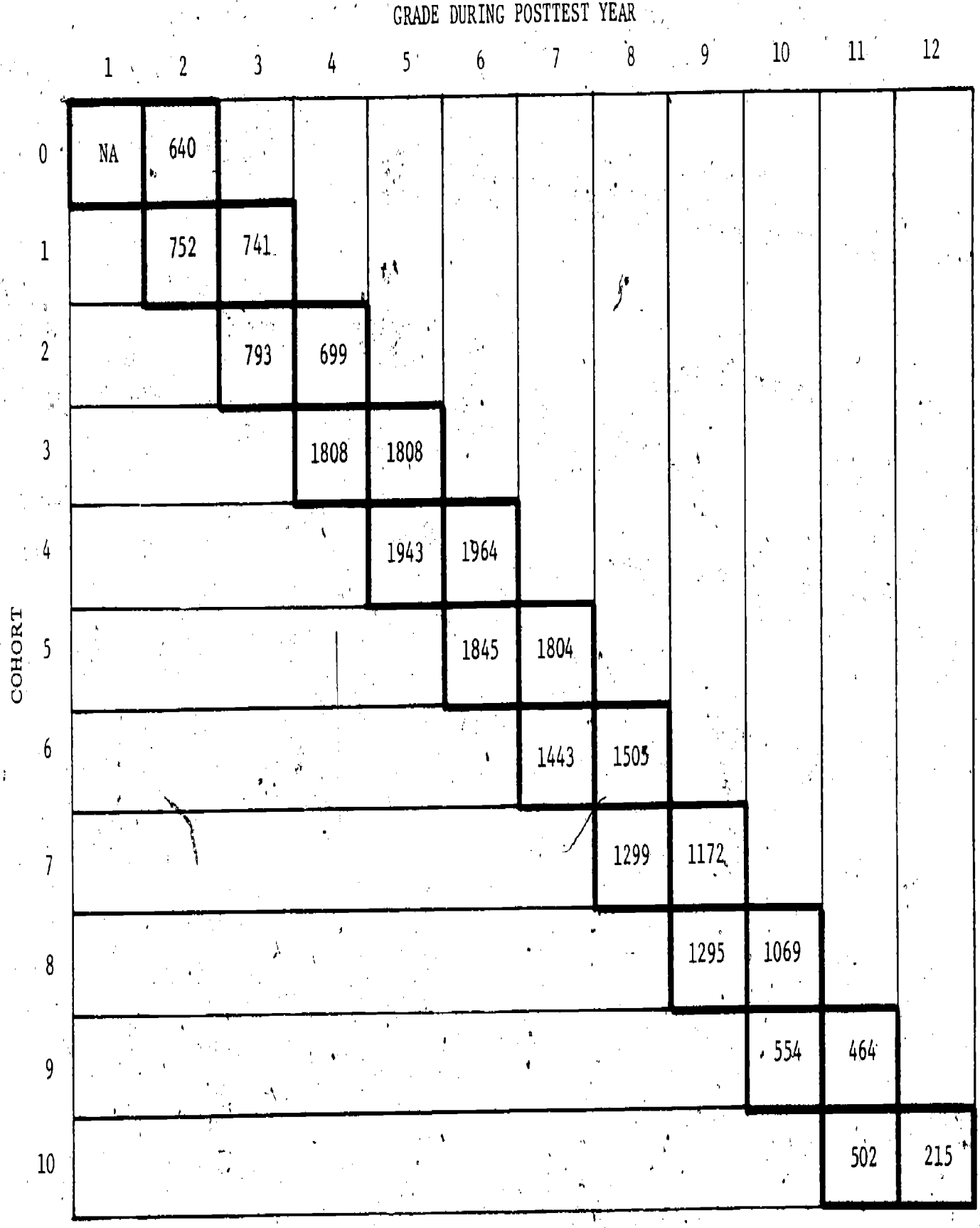


TABLE D-4

Pretest Means and Standard Deviations - Arithmetic Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	93.20* 12.86										
1		93.83* 12.12	319.15 37.43									
2			106.12* 10.20	369.69 43.47								
3				374.60 37.93	406.92 38.41							
4					404.69 37.68	449.44 56.06						
5						448.63 56.88	482.46 57.52					
6							487.46 57.44	517.17 66.72				
7								514.57 67.04	560.62 64.66			
8									552.94 60.22	609.38 73.20		
9										606.62 74.12	643.54 68.96	
10											647.48 70.32	676.94 66.95

*Based on the CTMM.

TABLE D-5

Socioeconomic Status (SES) Means and Standard Deviations -
Arithmetic Total Score Analyses

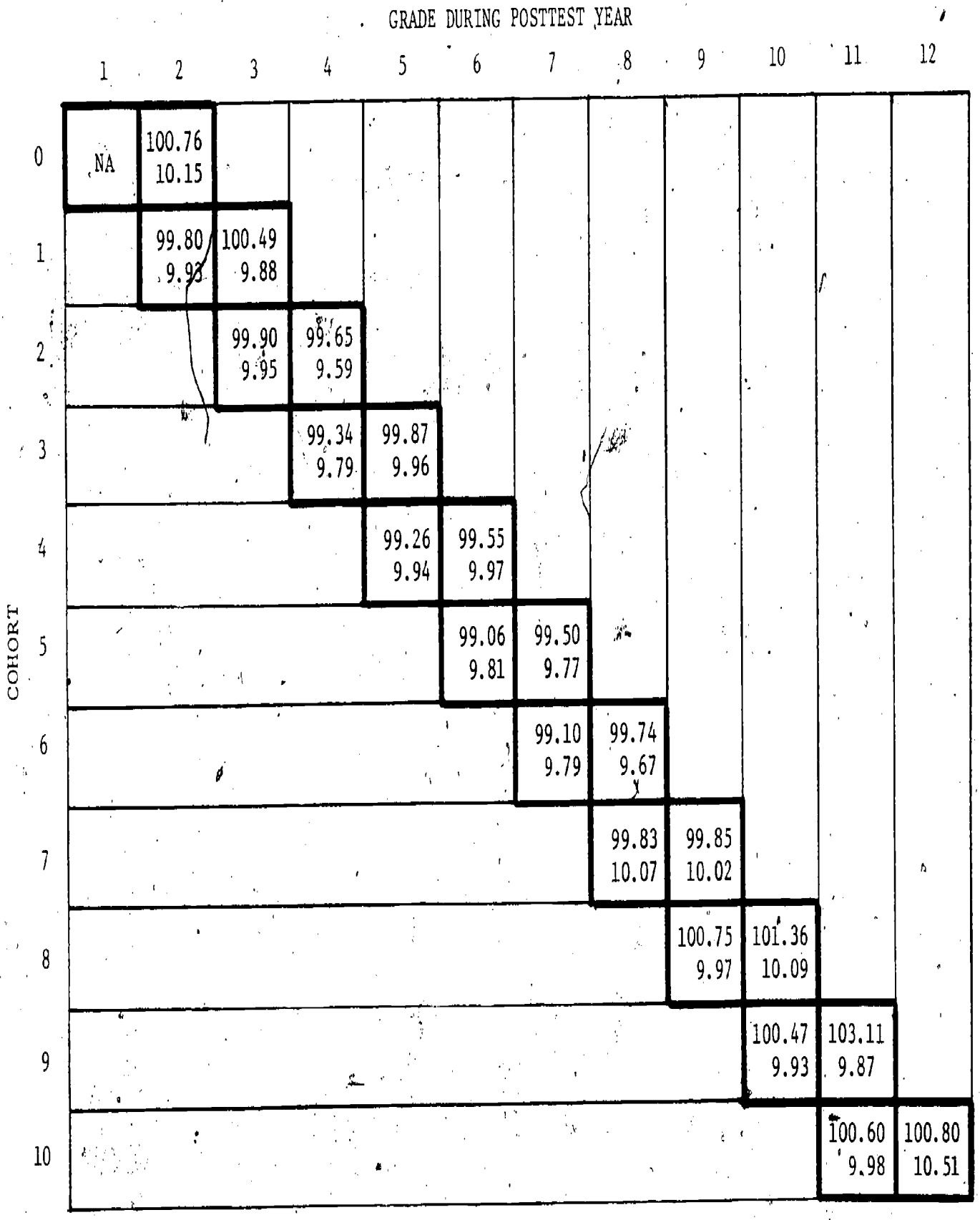


TABLE D-6

Level of Innovation Means and Standard Deviations - Arithmetic Total Score Analyses

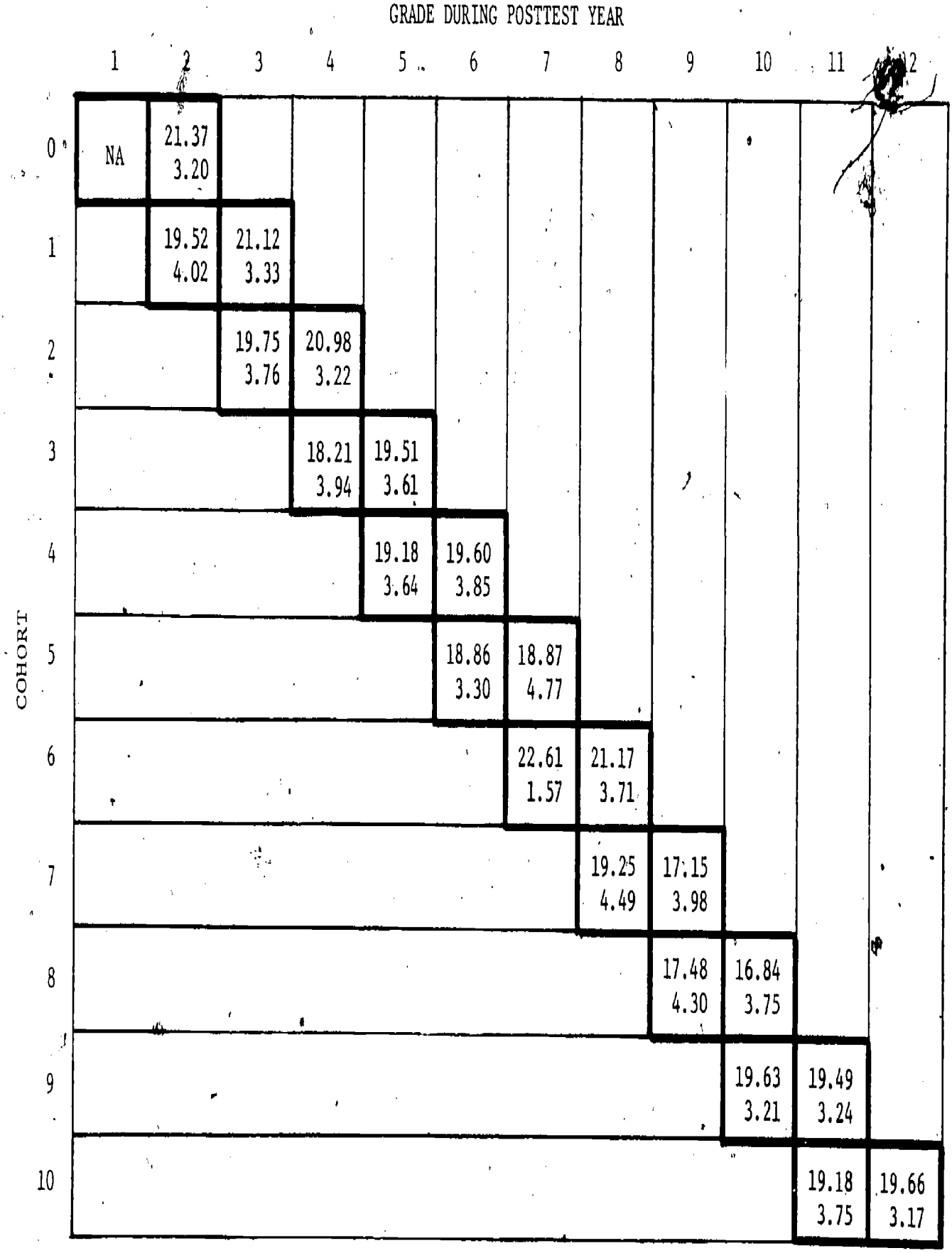


TABLE D-7

Number of Minutes per Day Means and Standard Deviations -
Arithmetic Total Score Analyses

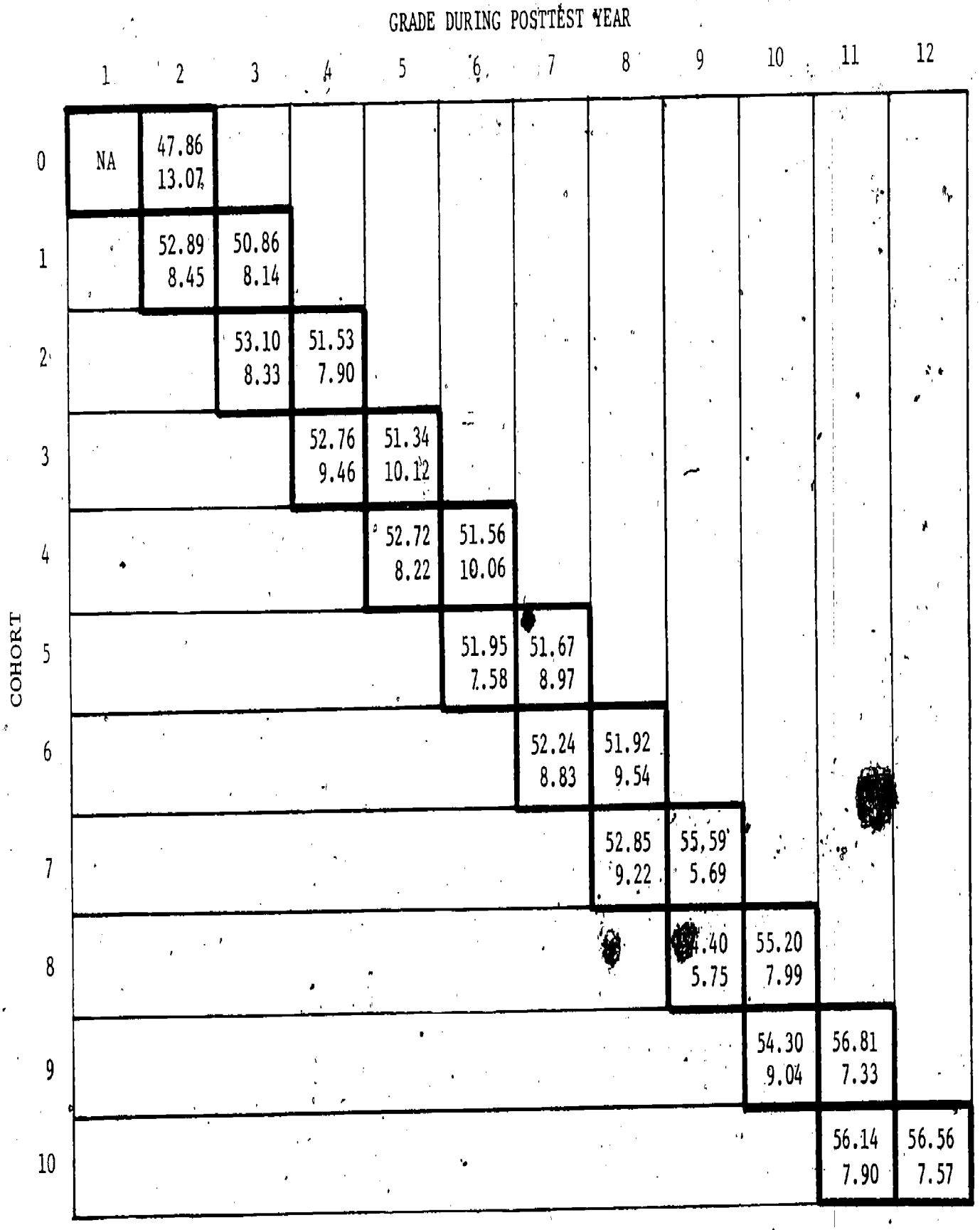


TABLE D-8

Teaching Qualifications Means and Standard Deviations - Arithmetic Total Score Analyses

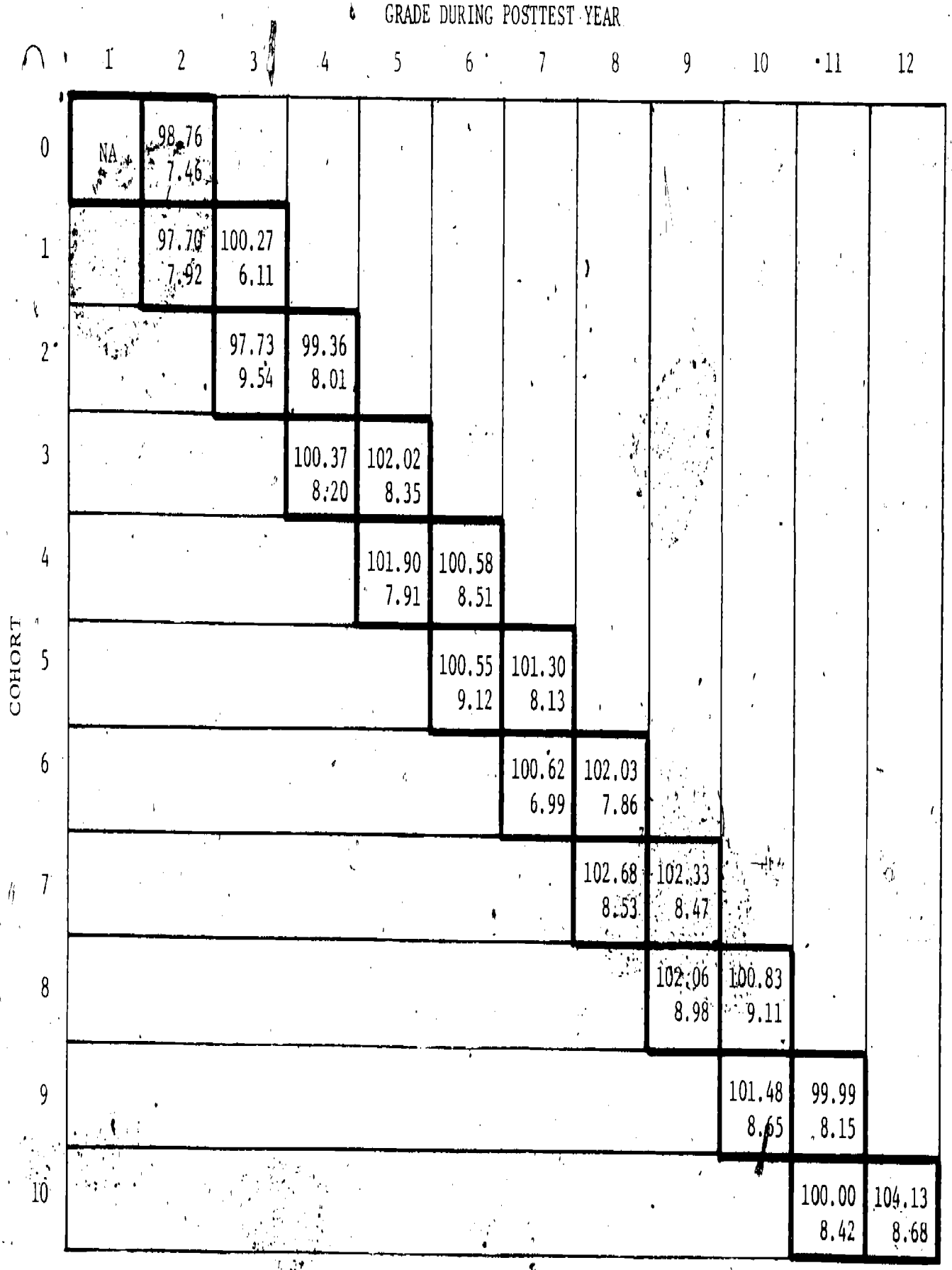


TABLE D-9

Uniquenesses and Multiple Correlations (Squared)
for Pretest - Arithmetic Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	.151 .159										
1		.106 .159	.461 .564									
2			.225 .263	.528 .599								
3				.491 .598	.492 .633							
4					.537 .629	.532 .666						
5						.602 .666	.599 .681					
6							.565 .663	.584 .731				
7								.649 .715	.523 .701			
8									.551 .670	.642 .707		
9										.663 .762	.658 .678	
10											.666 .693	.554 .564

6-P.

TABLE D-10

Uniquenesses and Multiple Correlations (Squared)
for SES - Arithmetic Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	.NA	.004 .023										
1		.083 .118	.000 .055									
2			.044 .069	.007 .070								
3				.012 .114	.012 .151							
4					.004 .090	.004 .097						
5						.004 .072	.001 .071					
6							.004 .084	.003 .085				
7								.001 .058	.014 .085			
8									.002 .043	.003 .049		
9										.002 .068	.002 .025	
10											.001 .003	.000 .000

D-10

TABLE D-11

Uniquenesses and Multiple Correlations (Squared) for Level of Innovation and Number of Minutes per Day - Arithmetic Total Score Analyses

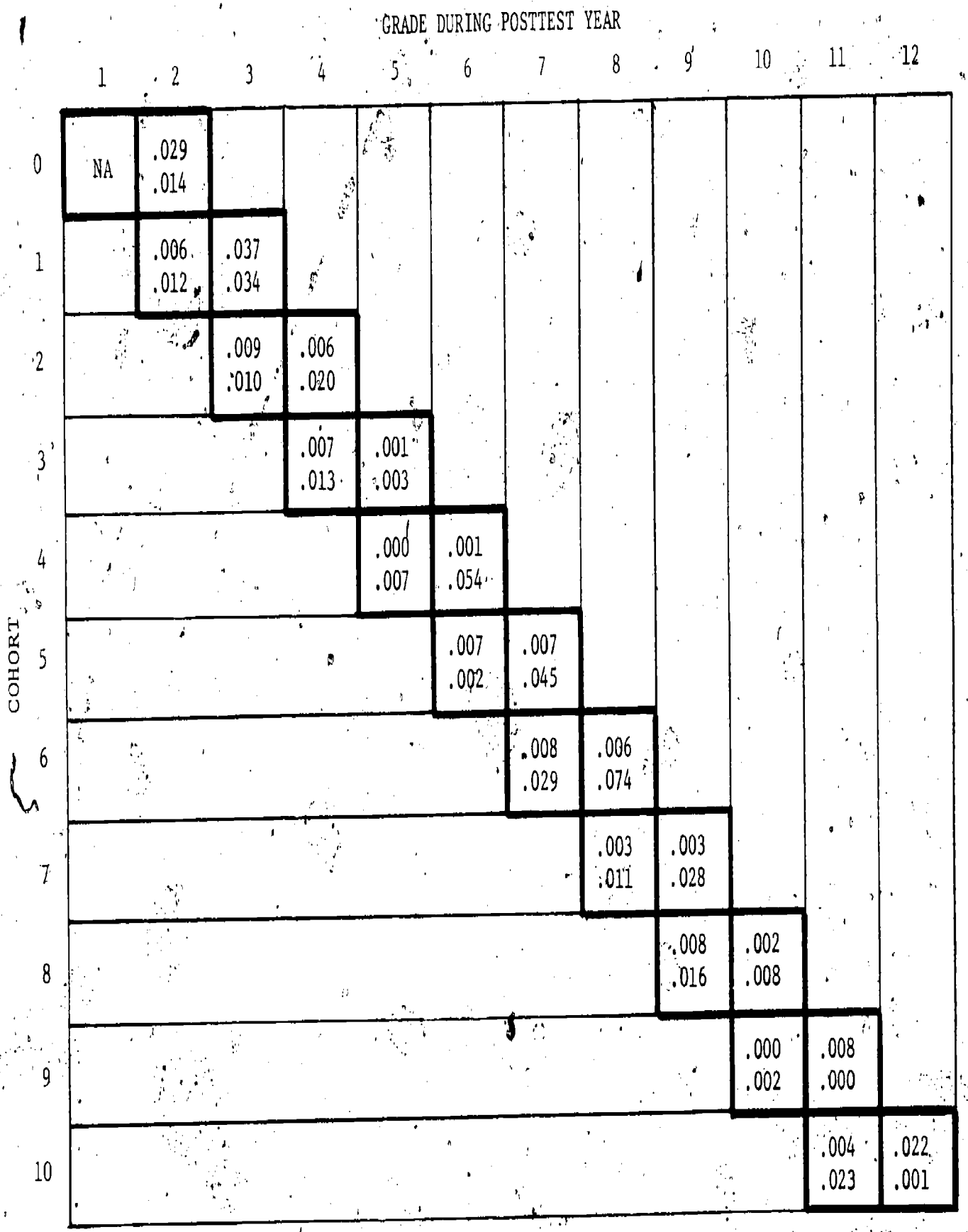


TABLE D-12

Uniquenesses and Multiple Correlations (Squared)
for Teaching Qualifications - Arithmetic Total Score Analyses

COHORT	GRADE DURING POSTTEST YEAR 1	2	3	4	5	6	7	8	9	10	11	12
0	NA	.001 .009										
1		.000 .000	.004 .034									
2			.005 .004	.008 .013								
3				.008 .022	.000 .003							
4					.000 .002	.000 .003						
5						.001 .000	.002 .014					
6							.003 .000	.000 .015				
7								.000 .005	.001 .061			
8									.007 .081	.000 .009		
9										.003 .020	.000 .008	
10											.004 .003	.001 .033

TABLE D-13

Numbers of Consistently Overachieving and Underachieving Students Identified - Arithmetic Total Score Analyses

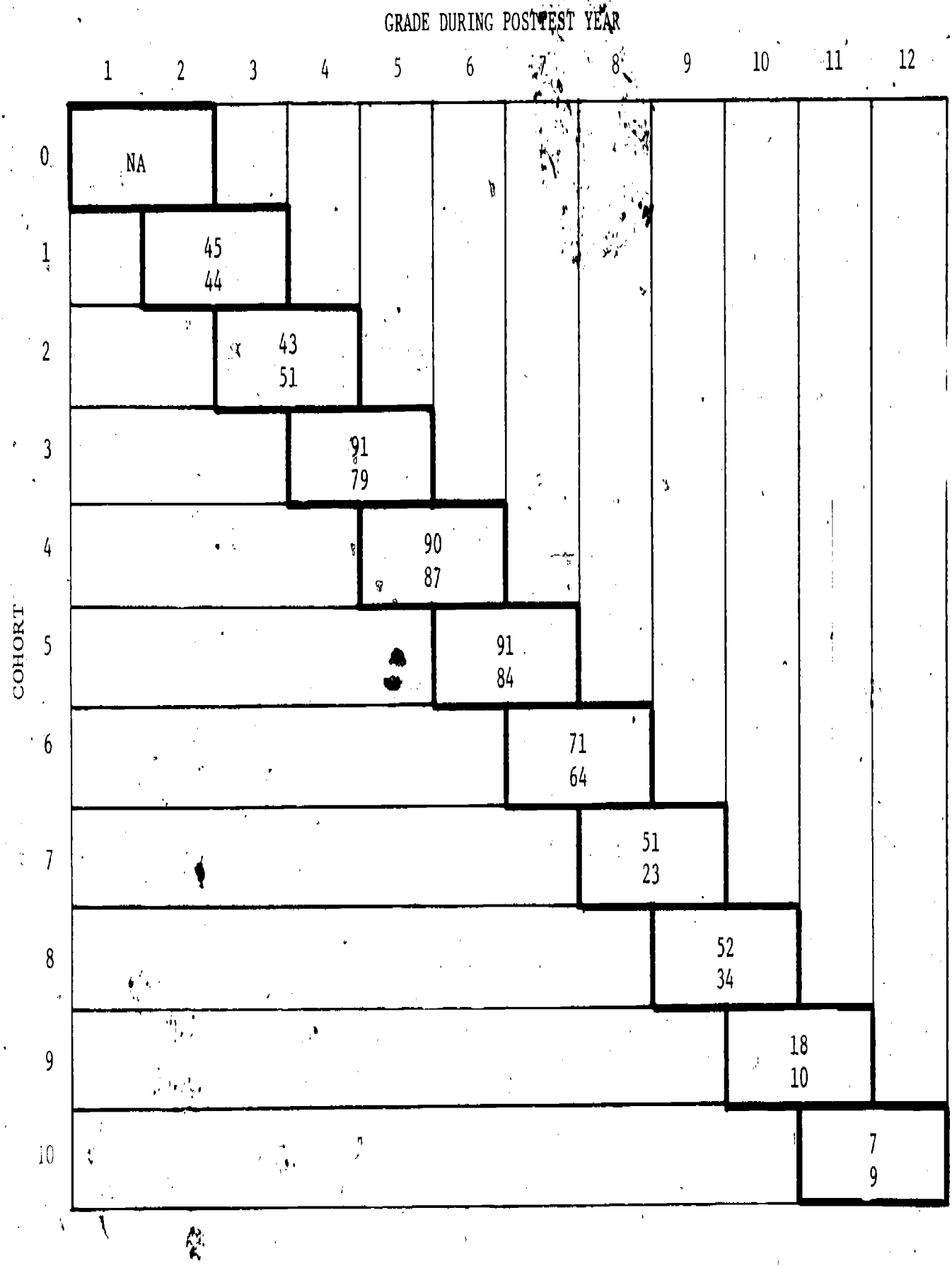


TABLE D-14

Point-Biserial Correlations Between Student Outlier Group Membership and Pretest (Year 1) - Arithmetic Total Score Analyses

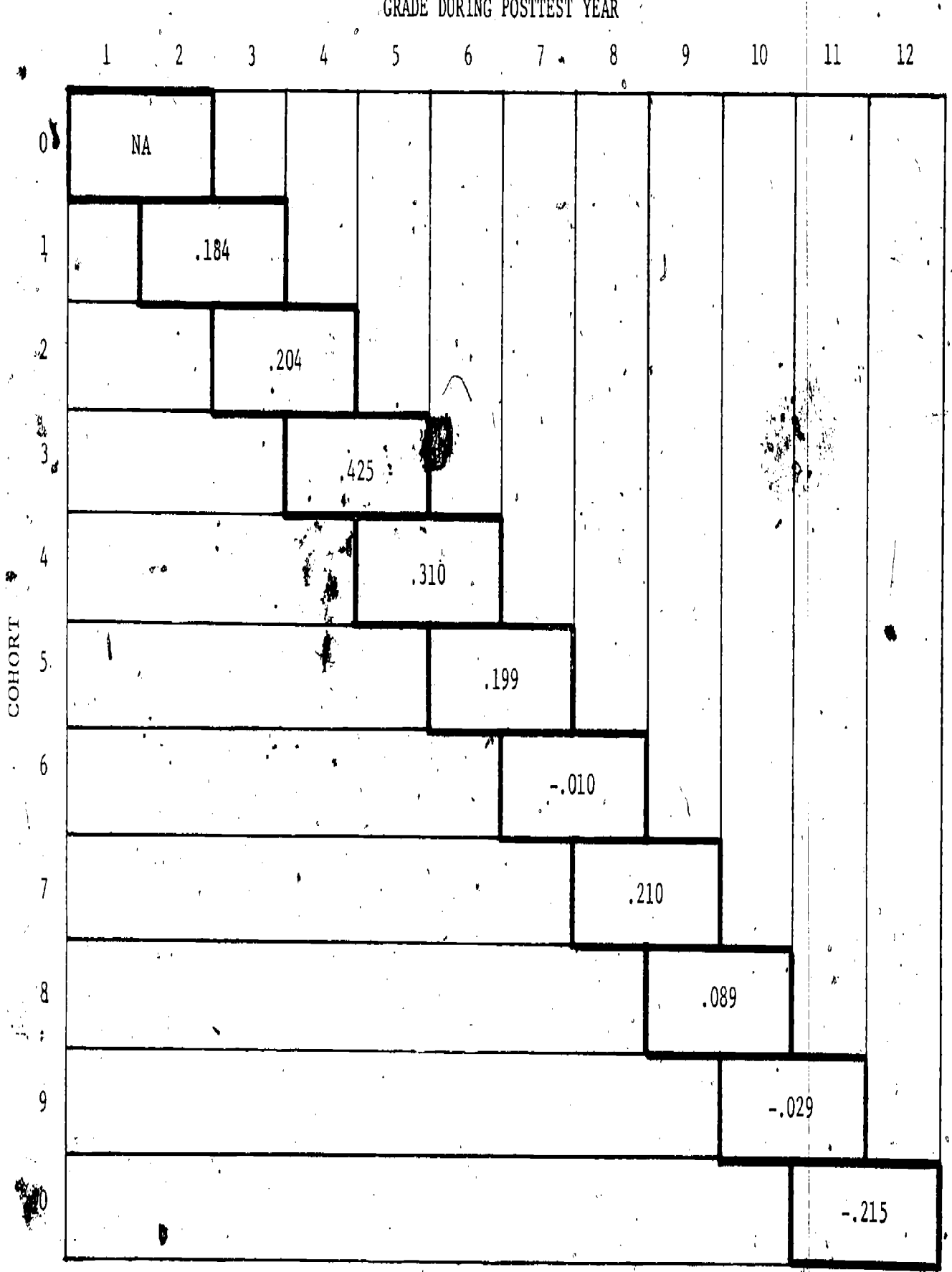


Table D-15

Point-Biserial Correlations Between Student Outlier Group Membership and SES - Arithmetic Total Score Analyses

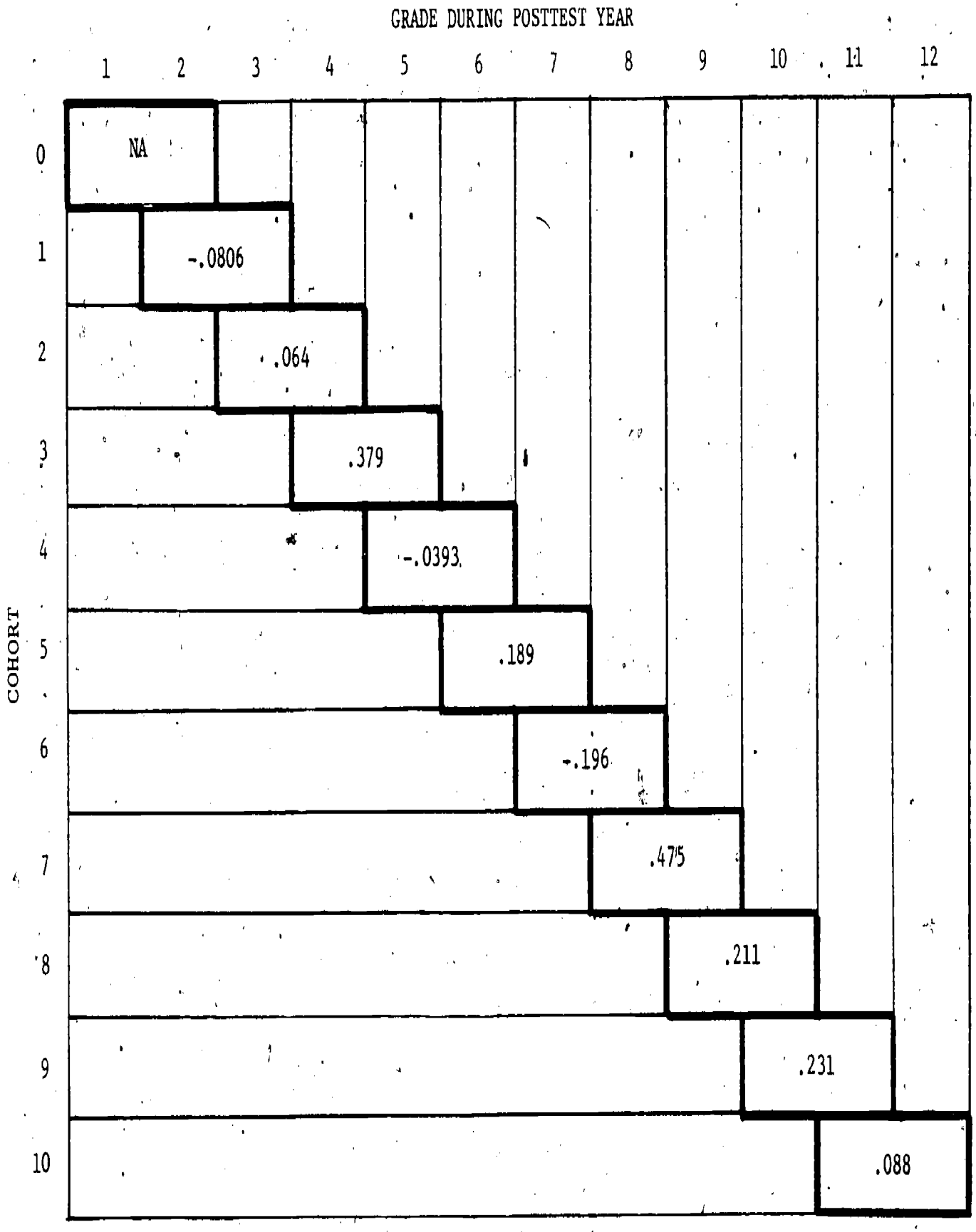


Table D-16

Point-Biserial Correlations Between Student Outlier Group Membership and Year 2 and Year 3 Test Score - Arithmetic Total Score Analyses

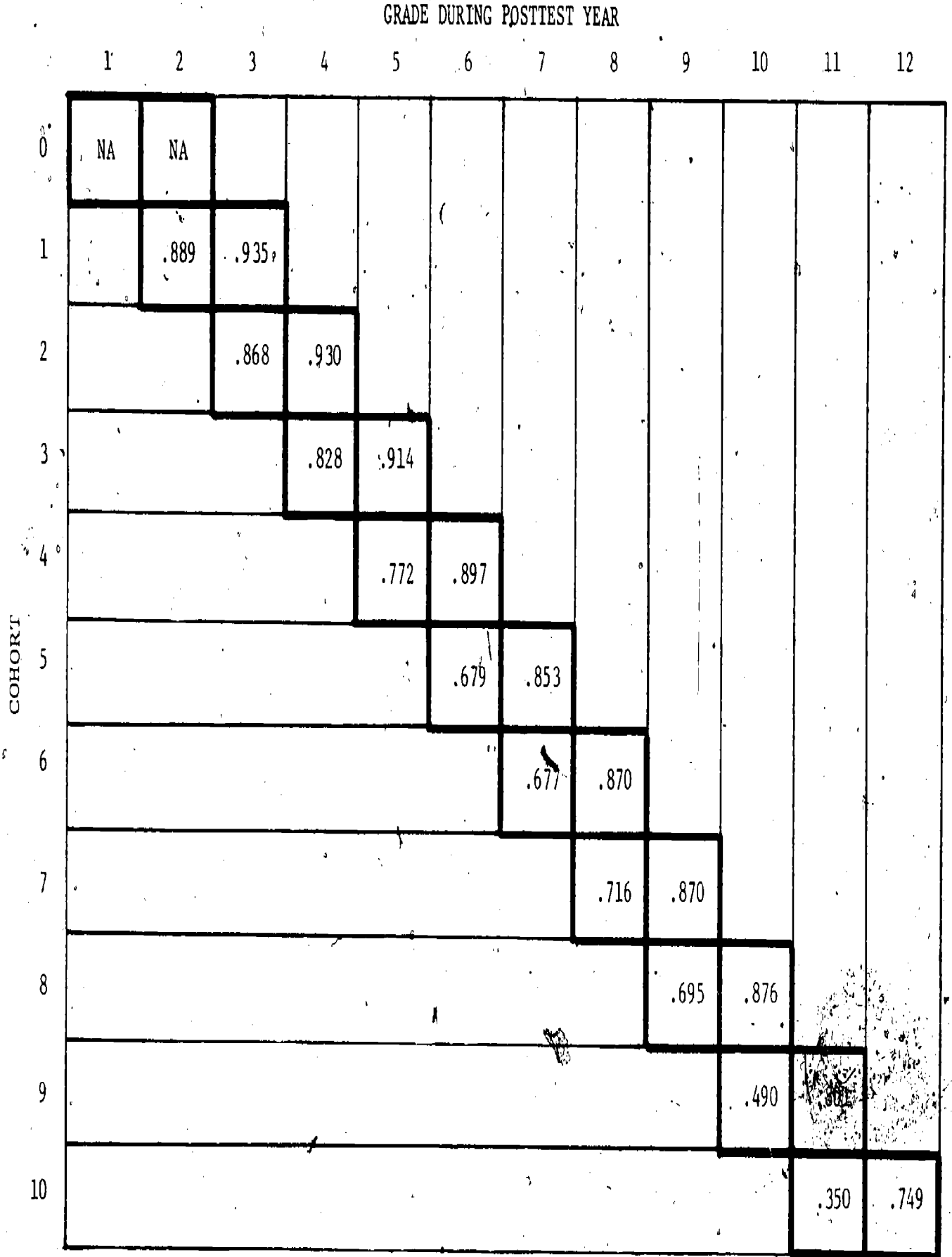


Table D-17

Point-Biserial Correlations Between Student Outlier Group Membership and Level of Innovation - Arithmetic Total Score Analyses

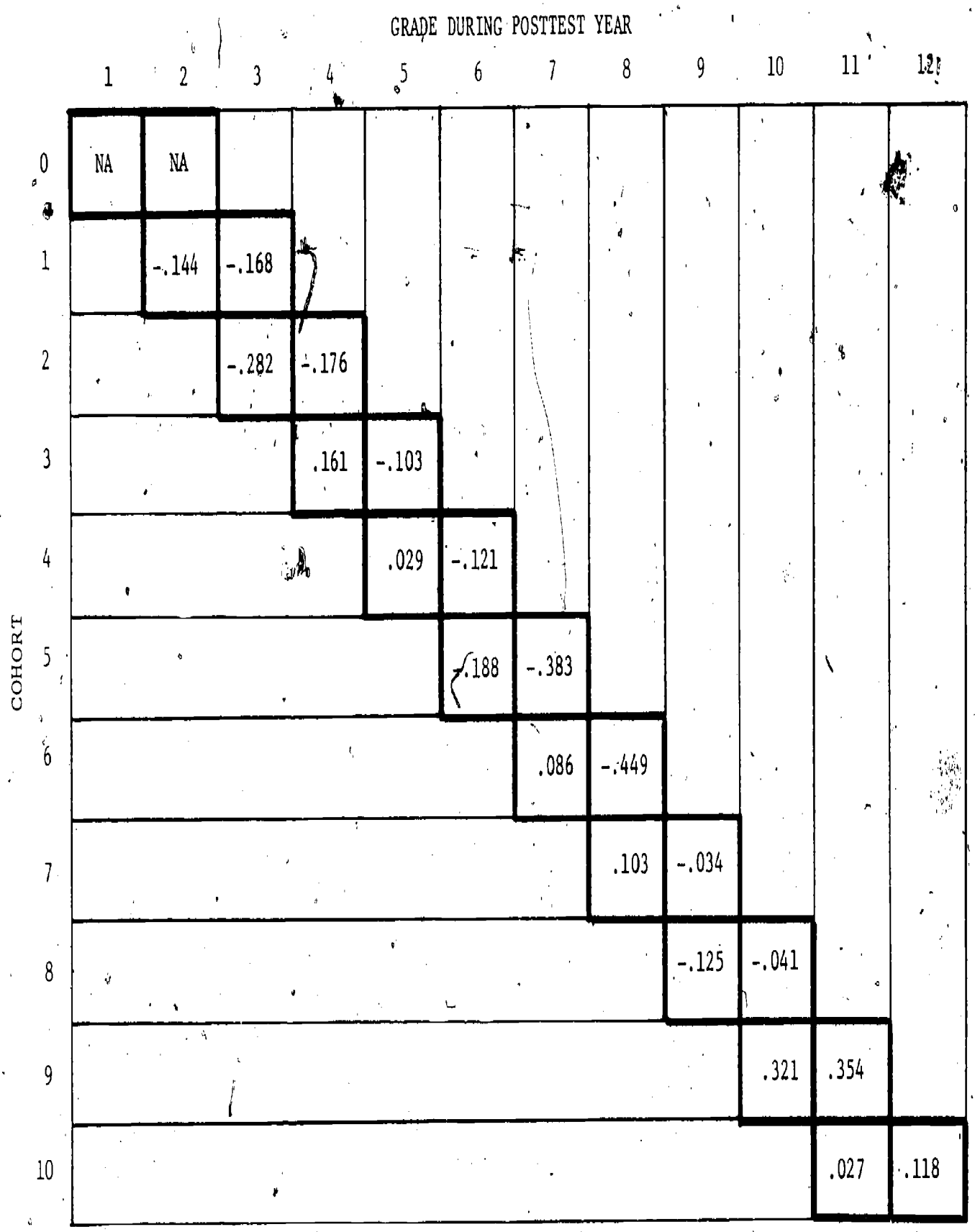


Table D-18

Point-Biserial Correlations Between Student Outlier Group Membership and Number of Minutes per Day - Arithmetic Total Score Analyses

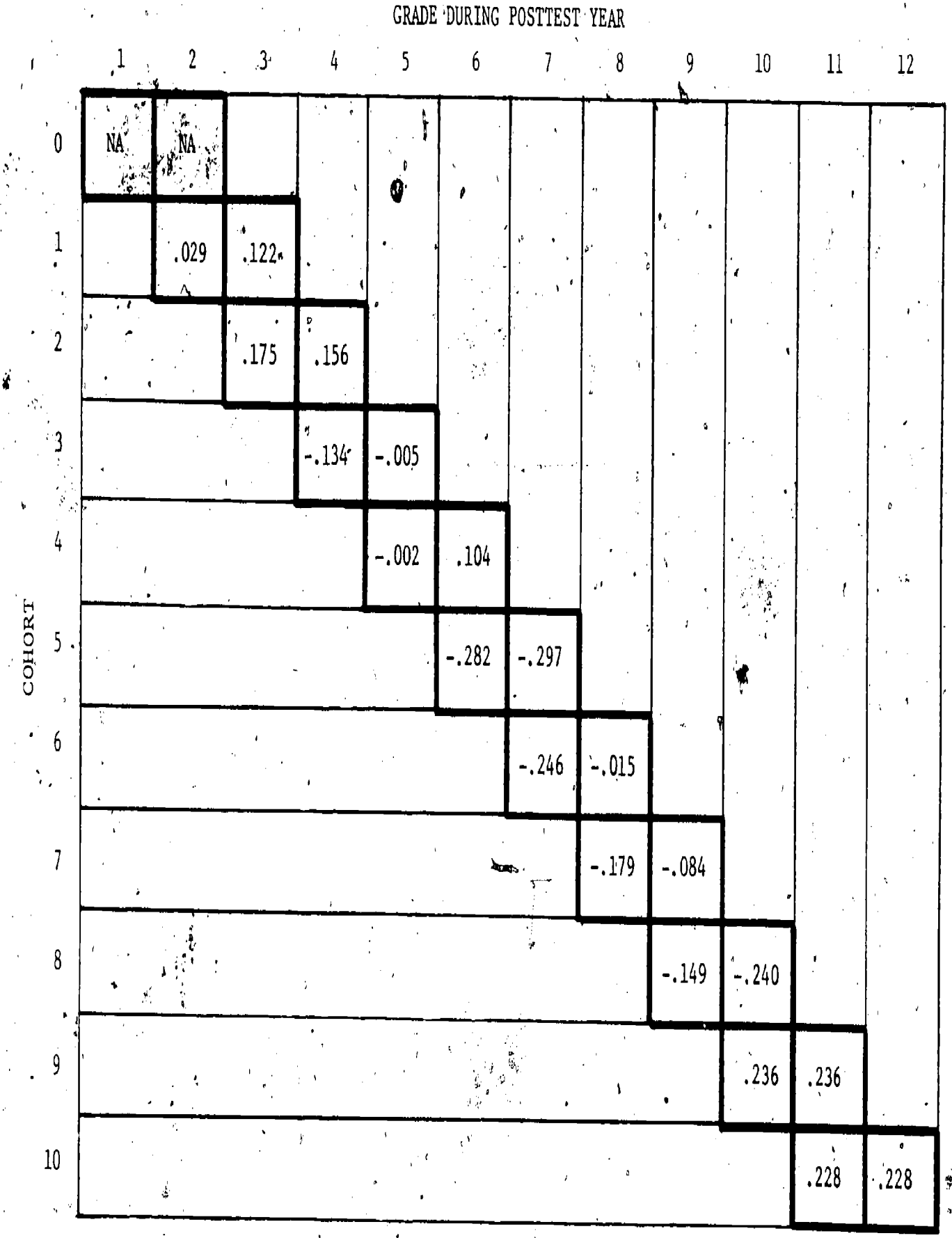
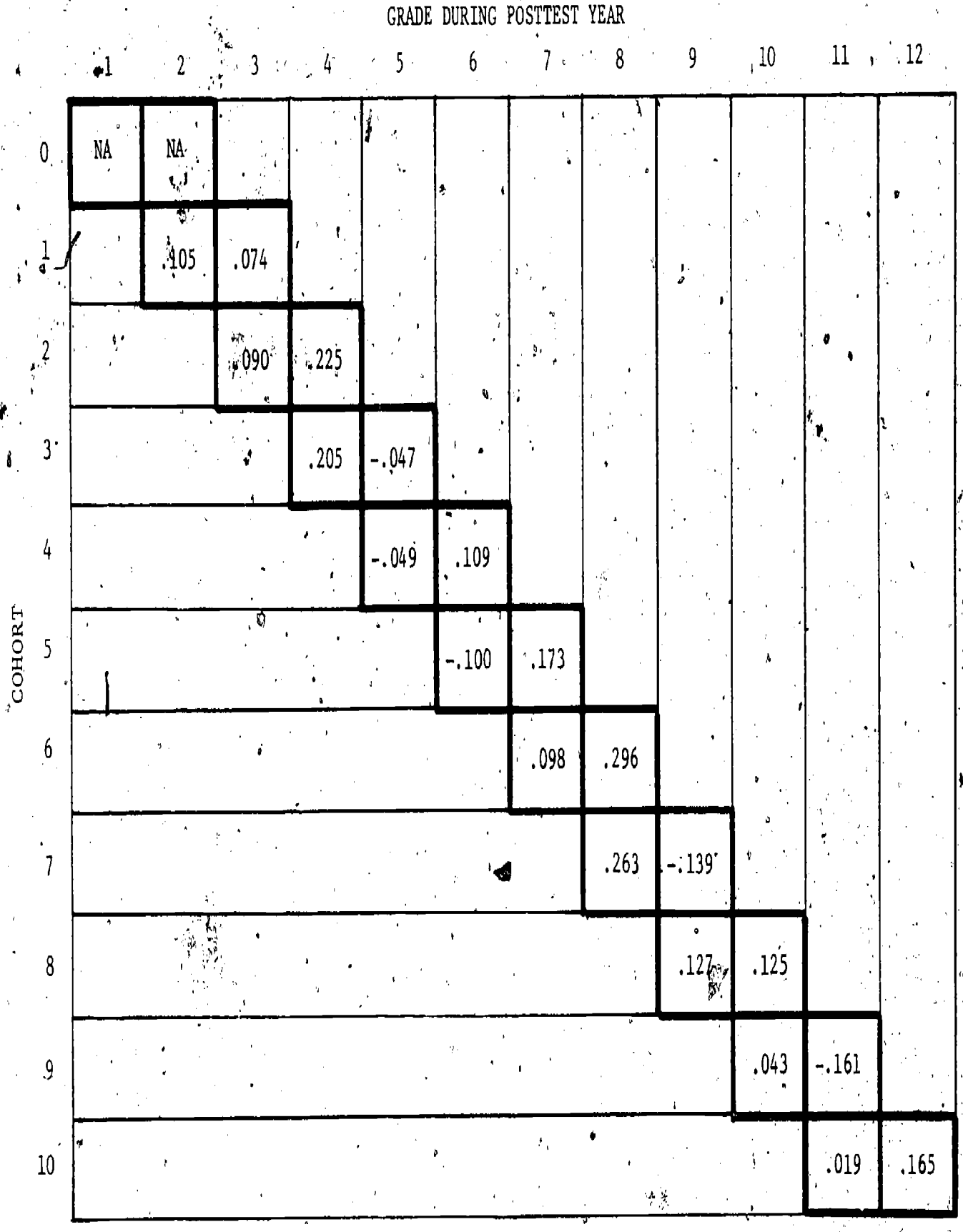


Table D-19

Point-Biserial Correlations Between Student Outlier Group Membership and Teaching Qualifications - Arithmetic Total Score Analyses



61-P